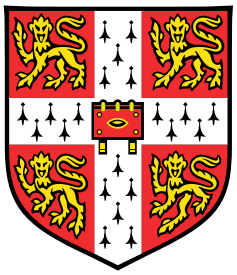
Understanding the Dynamics of Embryonic Stem Cell Differentiation



Stanley E. Strawbridge

Department of Biochemistry
Wellcome Trust-MRC Cambridge Stem Cell Institute
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

Girton College Michaelmas 2018

Understanding the Dynamics of Embryonic Stem Cell Differentiation

Stanley E. Strawbridge

The two defining features of mouse embryonic stem (ES) cells are self-renewal and naïve pluripotency, the ability to give rise to all cell lineages in the adult body. In addition to being a unique and interesting cell type, pluripotent ES cells have demonstrated their potential for continued advancements in biomedical science. Currently, there is an improved understanding in the chemical signals and the gene regulatory network responsible for the maintenance of ES cells in the naïve pluripotent state. However, less is understood about how ES cells exit pluripotency. My main aim is to study the dynamics and the factors affecting the irreversible exit from pluripotency.

Expression of the reporter Rex1-GFPd2, which is inactivated upon exit from naïve pluripotency, was analyzed by quantitative long-term single-cell imaging over many generations. This technique allowed chemical, physical, and genealogical information to be recorded during the transition to exit. Culture conditions that provided homogeneous populations were used in all assays and these data were validated against bulk-culture data where appropriate. Changes in real-time cell behavior were seen in cell-cell contact, motility, and cell-cycle duration. Undifferentiated ES cells form tightly joined colonies, with cells that exhibit low motility and a constant cell-cycle duration. Exit is associated with increasing cell motility, decreased cell-cell contact, and an acceleration in cell proliferation.

The onset of exit is associated with a sudden and irreversible inactivation of the Rex1-GFPd2 reporter. This inactivation is asynchronous, as it occurs at different times and in different generations during ES cell differentiation. However, examination of daughter cells generated from the same mother revealed a high level of synchronicity. Further investigation revealed that high levels of correlation in cell-cycle duration and Rex1-GFPd2 expression

exist between differentiating sister and cousin cells, providing strong evidence that cell potency is inherited symmetrically in cell divisions during exit *in vitro*.

How cells change fate is a fundamental question in developmental biology. Knowing the cellular dynamics during the transition out of naïve pluripotency is important for harnessing the potential of ES cells and understanding how cell fate decisions are made during embryonic development. The quantification of the timing of exit from naïve pluripotency coupled with identifiable changes in cellular behaviors, such as motility, cell size, and cell-cycle duration, enhances the understanding of how cell fate changes are regulated during directed differentiation.

In dedication to my loving grandparents,
Whom I have been so lucky to have in my life for so long.
Joseph Dale & Gladys Elaine Strawbridge
Eugene Carol & June Violet Grden

Declaration

I hereby declare that this dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. It does not exceed the prescribed word limit of 60,000 words.

Stanley E. Strawbridge
Michaelmas 2018

Acknowledgements

Thank you to my supervisors, Austin Smith, Graziano Martello, and Hillel Kugler. Thank you for the opportunity to study at Cambridge and for bringing me into the exciting world of stem cell biology. It has been a remarkable experience to be surrounded by so many bright and engaging people. Thank you for the time and energy you have expended fostering a project that crosses disciplinary boundaries, and for recognizing the value in performing such work. Finally, thank you for your guidance, support, and understanding throughout my PhD studies.

Thank you to the SCI facilities team. Peter Humphreys for training in all things microscopy and for the many discussions on imaging processing and cell tracking. Andy Riddell for enabling my independence in flow cytometry and the excellent talks about laser physics. And Paul Sumption, Paul Browne, and Paul Barrow for letting/tolerating me bypass/bypassing the IT ticketing service to talk computing.

Thank you to Guy Blanchard and Richard Adams for tailoring the tracking software Tracer and oTracks to fit my needs. Without these tools, I do not believe it would have been possible to track the thousands of hours of footage that I have.

Thank you to Jenny Nichols, Thorsten Boroviak, Sissy Wamaitha, and Fiona Price-Kuehne for being some of the the few who will ever read this document and for the much appreciated comments and feedback.

Thank you to Jo Jack and Austin Smith for printing, binding, and submitting this dissertation in my absence, so that I could be home with my family for Christmas.

Thank you to Microsoft Research for funding this project.

Thank you to all those who provided moral support:

To Thorsten Boroviak, Kevin Chalut, Paul Bertone, and Jenny Nichols. Thank you for all the thoughtful discussions about stem cell biology, development, and the role that mathematics can play in the biological world.

To all of my friends that I have acquired from St Catharine's College: Fiona McTigue (*née.* Docherty), Joshua McTigue, Amy Riddell, Sarah Connors, Aleksandra Kotwica, Will Bernard, and Matthew Sale. Thank for being my Cambridge family.

To FWC, Liam Simmonds, Ruby Newton, and Sissy Wamaitha. Thank you for all the adventures and for seeing me through the last years of my PhD. Here's to many more.

To Philippa Russell and all the people I have met through my various public engagement activities. Thank you for being uplifting and for always bringing perspective to my work.

To all in the Smith Group. I could not have asked for a more caring and intelligent set of colleagues and friends. Thank you to Carla Mulas, Tüzer Kalkan, Maria Rostovskaya, Martin Leeb, Rosalind Drummond, and Ge Guo for showing me the ropes of wet-lab stem cell biology and Amy Li and Masaki Kinoshita for all the technical advice in constructing my cell lines. The Late Shift, Yaoyao Chen, Agata Kurowski, and Yasu Takashima, thank you for keeping me company during time-points in the wee hours of the morning. And James Clarke, Sam Myers, and Nicholas Bredenkamp for the many pints and pub visits.

And to my family, Steve, Susan, and Stevie. Thank you for the love and support during my career as a professional student, from West Virginia to Oxford, and now finally Cambridge. I would not have been able to do any of this without you.

Table of contents

List of figures

List of tables

Abbreviations

2i ES cells	Either Rex1-GFPd2 or Rex1-GFPd2/Gap43 ES cells cultured in 2i.
2i	Two Inhibitors, PD03 + Chiron
APC	Adenomatosis Polyposis Coli Protein
Bam	Bag-of-Marbles
Bmp4	Bone Morphogenic Protein 4
BmprI/II	Bmp Receptor I/II
CDK	Cyclin Dependent Kinase
CK1α	Casein Kinase I α
Chiron	Chiron99021, a specific inhibitor of GSK3
Cldn7	Claudin 7
Col18a	Collagen Type XVIII Alpha 1 Chain
DNA	Deoxyribonucleic Acid
Dpp	Decapentaplegic
Dusp9	Dual-Specificity Phosphatase 9
EC cell	Embryonal Carcinoma Cell
ECM	Extra Cellular Matrix
ES Cell	Embryonic Stem Cell
Erk	Extracellular Signal-Regulated Kinase

Esrrb	Estrogen Related Receptor Beta
FBS	Fetal Bovine Serum
Fgf4	Fibroblast Growth Factor 4
Fgfr	Fgf Receptor
Flcn	Folliculin
Fn1	Fibronectin 1
Fnip1/2	Flcn Interacting Protein 1/2
GFP	Green Fluorescent Protein
GSC	Germline Stem Cell
Gbx2	Gastrulation Brain Homeobox 2
Gsk3	Glycogen Synthase Kinase 3
HMG	High Mobility Group
ICM	Inner Cell Mass
Icam1	Intercellular Adhesion Molecule 1
Id1/2/3	Inhibitor of Differentiation 1/2/3
Jak	Janus kinase
Klf2/4	Krüppel-Like Factor 2/4
LPNB	Lif & PD03 in N2B27
LPS ES cells	Rex1-GFPd2 ES cells cultured in Lif & PD03 in Serum.
LPS	Lif & PD03 in Serum
Lam(a1/c1)	Laminin Subunit Alpha1/Gamma1
Lifr	Lif Receptor
Lif	Leukemia Inhibitory Factor

Lrp	Lipoprotein Receptor-Related Protein
MAPK	Mitogen-Activated Protein Kinase
Mek	Mitogen/Extracellular Signal-Regulated Kinase
Myh10	Myosin Heavy Chain 10, aka Non-Muscle Myosin Heavy Chain IIb
Myh13	Myosin Heavy Chain 13
N2B27 cells	Either Rex1-GFPd2 or Rex1-GFPd2/Gap43-mCherry ES cells exiting in N2B27.
NuRD	Nucleosome Remodeling Deacetylase Complex
Oct4	Octamer-Binding Transcription Factor 4
PBS	Phosphate Buffered Saline
PD03	PD0325901, a specific inhibitor of Mek
PEST	Proline, Glutamic Acid, Serine, and Threonine
Pecam1	Platelet And Endothelial Cell Adhesion Molecule 1
RB	Retinoblastoma Protein
RNA	Ribonucleic Acid
RTK	Receptor Tyrosine Kinase
Rex1	also called ZFP42
Sall4	Sal-Like Protein 4
Smad1/4/5	portmanteau of *C. elegans* Sma (small body size) and *D. melanogaster* Mad (mothers against decapentaplegic) 1/4/5
Sox2	SRY-related HMG Box 2
Stat3	Signal Transducer and Activator of Transcription 3
Tbx3	T-Box 3
Tcf3	Transcription Factor 3, also known as

Tfcp2l1	Transcription Factor CP2 Like 1
Tfe3	Transcription Factor E3
Tsc1/2	Tuberous Sclerosis 1
UTR	Untranslated Region
Upd	Unpaired
WT 2i ES Cells	E14 IVC ES cells with the tracking reporter cultured in 2i.
Wnt	portmanteau of *D. melanogaster* Wg (wingless) and formerly named *M. musculus* int1
Zfp	Zinc Finger Protein
ex/em	Excitation and Emission
hPSC	Human Pluripotent Stem Cell
mCherry	Monomer Cherry (fluorescent protein)
serum cells	Rex1-GFPd2 ES cells exiting in Serum.

Chapter 1

Introduction

"WE SHALL NOT CEASE FROM EXPLORATION,
AND THE END OF ALL OUR EXPLORING
WILL BE TO ARRIVE WHERE WE STARTED
AND KNOW THE PLACE FOR THE FIRST TIME."
- T.S. ELIOT, 1943

1.1 Origins and Maintenance of Embryonic Stem Cells

Embryonic stem (ES) cells are derived from the epiblast of the inner cell mass (ICM) of the pre-implantation blastocyst (Fig. 1.1) (Evans and Kaufman, 1981; Martin, 1981; Nichols and Smith, 2012). The defining features of ES cells are self-renewal and pluripotency (Smith, 2006). Self-renewal is the ability to indefinitely produce unaltered daughter cells with the same differentiation potential (Smith, 2006). ES cells are also pluripotent, meaning they retain the ability to give rise to all cell types of the body, including the germ cells (Jaenisch and Young, 2008; Martello and Smith, 2014; Nichols and Smith, 2012; Smith, 2001; Young, 2011). Pluripotency can be demonstrated in a variety of ways, from the use of exogenous molecular signals to differentiate ES cells into all adult lineages *in vitro* (Jaenisch and Young, 2008; Smith, 2001) to *in vivo* blastocyst injection to generate germline competent chimeras (Bradley et al., 1984).

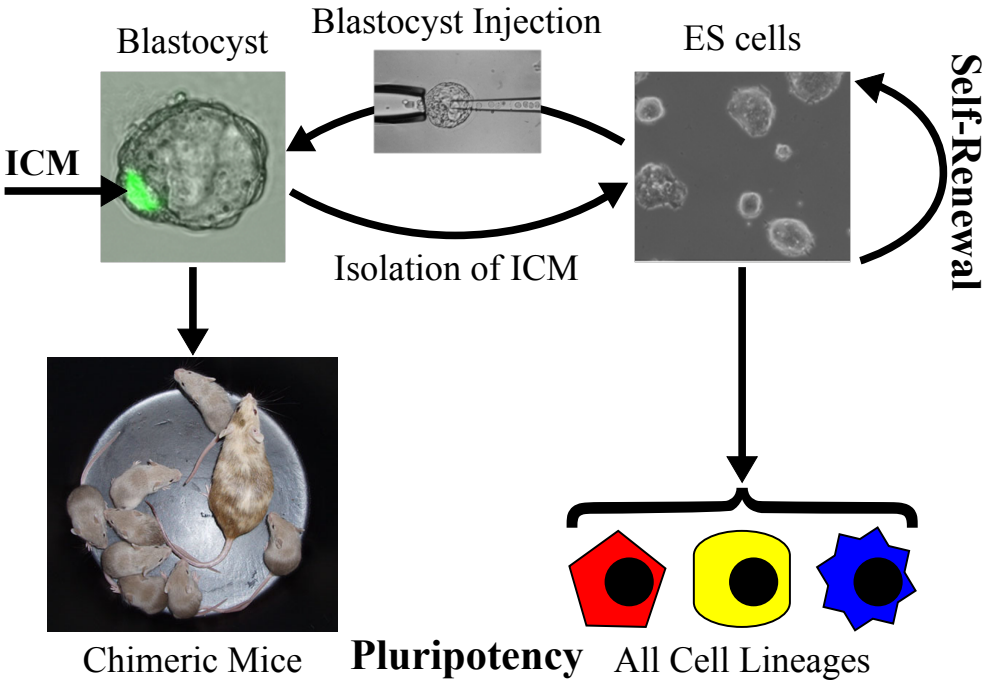


Fig. 1.1 ES cells are **pluripotent**, meaning they are able to give rise to all cell lineages of the adult organism, and are in a state of **self-renewal**, defined by the ability to generate at least one daughter equivalent to the mother with the same latent capacity for differentiation (Smith, 2006). Differentiation potential can be shown by *in vitro* differentiation or blastocyst injection to generate germline competent chimeras. (Unpublished images provided by J. Nichols).

The story of ES cells can be traced back to the testicular teratomas in the 129 mouse strain (Stevens and Little, 1954). Stevens, et al., described these teratomas as, "chaotic mixtures of many tissue types of various degrees of maturity," containing both embryonic and adult tissues from all three somatic lineages and germ cells (Stevens, 1958; Stevens and Little, 1954). They found "pluripotent embryonic cells" that gave rise to both rapidly differentiating cells and self-similar undifferentiated cells (Stevens and Little, 1954), the two hallmarks of pluripotent stem cells. Transplantation studies showed the necessity of these, now named, "embryonal" cells for successful teratoma propagation (Stevens, 1958).

Pluripotent embryonal carcinoma (EC) cells were isolated from teratomas and were shown *in vitro* to give rise to cell types from all germ layers (Kleinsmith and Pierce, 1964; Martin and Evans, 1975). EC cells were further shown to contribute to chimeras (Brinster, 1974), however, this was inefficient with low reproducibility and did not show evidence of germline transmission (reviewed by Smith, 2001). Clonal propagation and differentiation of EC cells were accomplished using teratoma conditioned medium supplemented with fetal calf serum (Kleinsmith and Pierce, 1964; Martin and Evans, 1975), but EC cells were found

to be more robust when co-cultured with mitoticaly inactivated embryonic fibroblasts, or "feeders" (Martin, 1980; Martin et al., 1977).

Although pluripotent cells were shown to exist in the pathology of a tumor, it was not known if there existed an instance of non-aberrant pluripotent cells? Teratoma grafts share similarity to normal mouse development by *in vivo* self-organization of the germ layers (Stevens, 1959). Reciprocally, grafts of the pre- and post-implantation embryo have the capacity to generate teratomas (Solter et al., 1970; Stevens, 1970). The pre-gastrulation epiblast was identified as the region containing these "normally" occurring pluripotent cells (Diwan and Stevens, 1976).

The first ES cell lines were derived from outgrowths of pre-implantation mouse embryo explants onto a feeder-layer (Evans and Kaufman, 1981; Martin, 1981). These cells had the capacity to self-renew and were shown to contribute to normal development by the generation of germline competent chimeras (Beddington and Robertson, 1989; Bradley et al., 1984). Similar to EC cultures using teratoma conditioned medium, ES cells can be propagated in medium conditioned by "feeders", mouse embryonic fibroblasts (Smith and Hooper, 1987). Leukemia inhibitory factor (Lif) is the trophic factor found in the feeder conditioned medium, and serum medium supplemented with Lif is sufficient to maintain a feeder-free ES cell culture (Smith et al., 1988; Williams et al., 1988).

While the necessary exogenous signals for the maintenance of ES cells were becoming more apparent, the endogenous factors and genetic basis of pluripotency was still unknown. Boyer et al., 2005, first proposed that the serum/Lif culture system was stabilizing a network of transcription factors that behaved as a self-sustaining circuit (Boyer et al., 2005; Loh et al., 2006; Rodda et al., 2005). The first three pluripotency factors identified to be members of this network were Oct4, Sox2, and Nanog (Avilion et al., 2003; Chambers et al., 2003, 2007; Masui et al., 2007; Mitsui et al., 2003; Nichols et al., 1998; Niwa et al., 2000; Schöler et al., 1990). Extensive studies have demonstrated their reciprocal nature in the regulation of one another and of other key pluripotency genes (Chew et al., 2005; Kuroda et al., 2005; Rodda et al., 2005).

In ES cells, Oct4 and Sox2 form the core of the pluripotency transcription factor network. Oct4 belongs to the Octamer class of the POU (Pit, Oct, and Unc) transcription factors (Chambers and Tomlinson, 2009; Nichols et al., 1998; Niwa et al., 2000; Schöler et al., 1990). Oct4 is necessary for establishing and maintaining pluripotency in the ICM (Nichols et al., 1998; Schöler et al., 1990). Silencing of Oct4 results in embryos which are able to implant but are only able to form the extra-embryonic tissue trophectoderm, the outer most and first specified tissue of the embryo (Niwa et al., 2000; Pesce and Schöler, 2001). Over expression

of Oct4 in ES cells results in differentiation towards the endodermal and mesodermal lineages (Nichols et al., 1998; Niwa et al., 2000).

Oct4 associates with and cooperatively binds DNA with Sox2 (Ambrosetti et al., 2000; Chambers and Smith, 2004; Pesce and Schöler, 2001). Sox2 is a transcription factor that possesses an HMG box and belongs to the SRY-related HMG Box (Sox) family (Avilion et al., 2003; Chambers and Tomlinson, 2009; Masui et al., 2007). Sox2 is expressed in parallel with Oct4 during early development prior to implantation (Avilion et al., 2003; Pesce and Schöler, 2001). Sox2 knock-outs are early embryo lethal, showing its necessary role in development (Avilion et al., 2003; Maruyama et al., 2005).

Nanog is a homeobox-containing transcription factor required for the acquisition and maintenance of pluripotency in the early embryo (Chambers et al., 2003; Mitsui et al., 2003; Silva et al., 2009). It has also been shown to be important later in development, during primordial germ cell specification (Chambers et al., 2003, 2007; Yamaguchi et al., 2005). So unlike Oct4 and Sox2, Nanog has a role distinct from the "housekeeping machinery of pluripotency" (Chambers et al., 2007). Over expression of Nanog is sufficient to maintain ES cells in the pluripotent state under pro-differentiation conditions (Chambers et al., 2003; Mitsui et al., 2003). Nanog knock-outs are prone to spontaneous differentiation towards primitive endoderm (Chambers, 2004; Mitsui et al., 2003), however there is evidence that, at low frequency, Nanog-null cells are able to contribute to the germline (Carter et al., 2014).

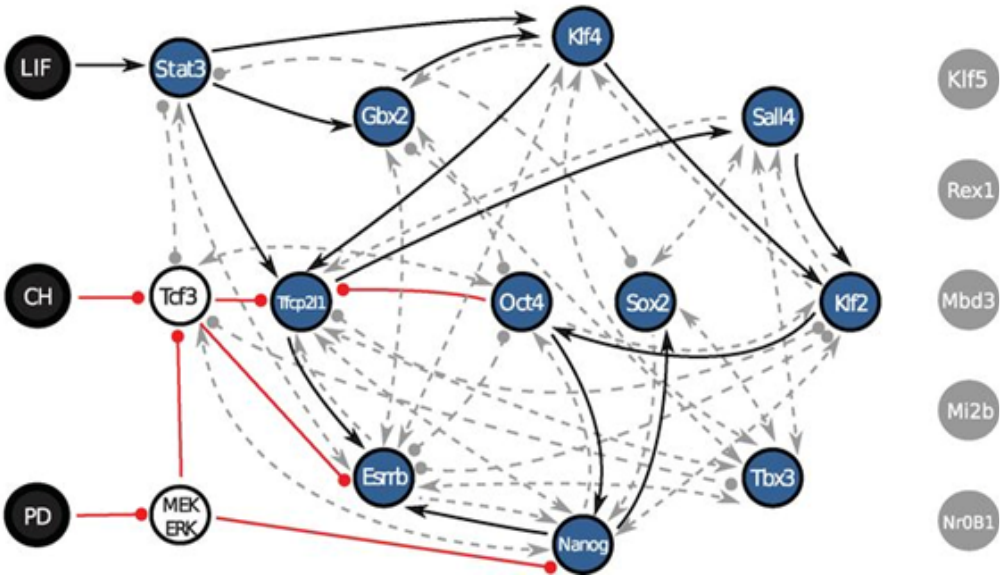


Fig. 1.2 Consensus topology of the naïve pluripotency transcription factor network. Oct4 and Sox2 sit at the heart of the transcription factor network which is surrounded by the naïve factors. Solid lines show the essential interaction which must occur to satisfy the empirical data. Black lines tipped with arrows indicate activation and red lines tipped with circles indicate repression. Blue circles are pluripotency factors, white circles are pro-differentiation factors, and black circles are exogenous signals used to stimulate the network (this will be discussed next). Many other factors and factor interactions have been implicated in this network, however all experimental constraints from this study can be satisfied without the the gray factors and interactions. (reproduced from Dunn et al., 2014).

Since Nanog was first sequenced, several other transcription factors have been identified to participate in the establishment of naïve pluripotency. These naïve factors include Esrrb, Gbx2, Klf2, Klf4, Sall4, Tbx3, and Tfcp2l1 (Cartwright et al., 2005; Dunn et al., 2014; Martello et al., 2013, 2012; Niwa et al., 1998, 2009; Yeo et al., 2014). The relationship of these factors is shown in Fig. 1.2. The topology of the consensus naïve pluripotency transcription factor network was determined using formal reasoning methods that enabled the synthesis of networks consistent with known experimental constraints (Dunn et al., 2014; Yordanov et al., 2016).

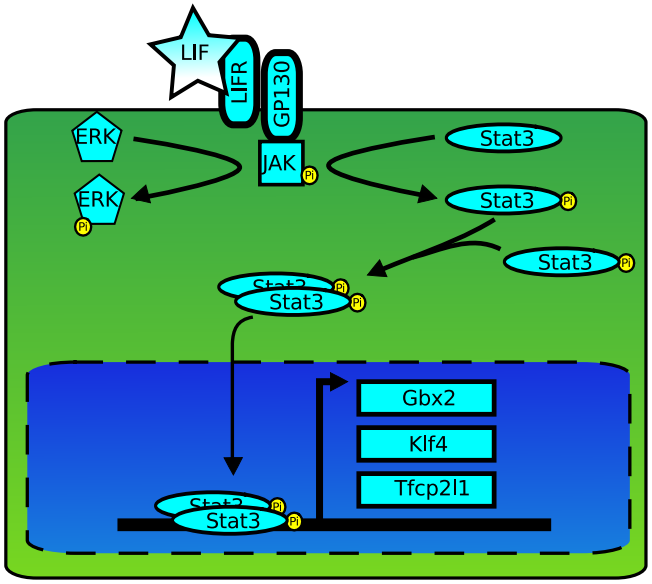


Fig. 1.3 Lif-Stat3 signaling in ES cells. Lif supports self-renewal in ES cells by activating the Jak-Stat3 signaling pathway. Binding of Lif by the Lif receptor (LIFR) induces dimerization with GP130 and subsequent activation of the Jak. Phosphorylation of Stat3 by Jak induces homodimerization of phospho-Stat3, which then translocates to the nucleus to activate transcription of naïve factors Gbx2, Klf4, and Tfcp2l1. However, Jak cross-talk also results in activation of pro-differentiation ERK. (adapted from Lodish, 2008; Quinlan, 2011).

Lif is now known to stimulate the Jak/Stat3 pathway, leading to the transcription of elements in the naïve pluripotency circuitry (Matsuda et al., 1999; Niwa et al., 1998). Fig. 1.3 shows relevant aspects of the Lif signaling pathway. Lif binds to Lif receptor (Lifr), causes dimerization with gp130, also know as IL6ST, which is associated with Jak (Burdon et al., 1999a; Lodish, 2008; Quinlan, 2011). Receptor dimerization induces the kinase activity of Jak, which phosphorylates the transcription factor Stat3 (Burdon et al., 1999a; Lodish, 2008; Quinlan, 2011). The phosphorylation of Stat3 induces homodimerizatoin and translocation to the nucleus. The Stat3 homodimer then activates transcription of Gbx2, Klf4, and Tfcp2l1 (Cartwright et al., 2005; Martello et al., 2013; Niwa et al., 1998, 2009). However, Lif signaling also results in the activation of pro-differentiation pathways via phosphorylation of Erk (Burdon et al., 1999a; Kunath et al., 2007; Niwa et al., 2009).

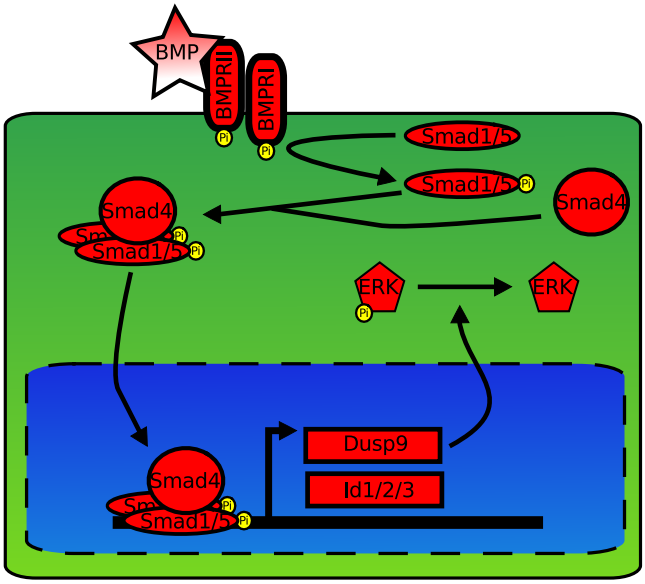


Fig. 1.4 Bmp4-Smad signaling in ES cells. Bmp4 is a factor found in serum which further stabilizes the pluripotency circuitry in serum/Lif culture conditions (Ying et al., 2003). Bmp binds the Bmp Receptors I and II which then phosphorylates Smad1/5. Dimerization of Smad 1/5 recruits Smad4 at which point the complex translocates to the nucleus to activate transcription of Id1/2/3, which block differentiation, and Dusp9, which dephosphorylates prodifferentaiton Erk. Dusp9 effectively balances Erk activation by Lif. (adapted from Lodish, 2008; Quinlan, 2011).

LIF is sufficient to maintain pluripotency in the presence of serum, yet, when ES cells are cultured in serum-free media with LIF, they spontaneously differentiate towards the neuroectoderm lineages (Ying et al., 2003). This indicates that the serum contains some factor(s) that regulate ES cell self-renewal. Bone morphogenic protein (Bmp) 4 has been shown to be the main factor in serum that counteracts differentiation by reducing Erk activity in serum/Lif cultures (Fig. 1.4, Ying et al., 2003). Bmp4 binds the Bmp receptor II (BmprII), causing heterodimerization with BmprI (Lodish, 2008; Quinlan, 2011). These receptors have inherent kinase activity that activates the Smad pathway via phosphorylation of Smad1 and Smad5 (Lodish, 2008; Quinlan, 2011). Dimerization of Smads1/5 induces association with Smad4, at which point the Smad 1/4/5 trimer translocates into the nucleus where it behaves as a transcription factor for dual-specificity phosphatase (Dusp) 9 and inhibitor of differentiation genes (Id) 1/2/3 (Li et al., 2012; Ruzinova and Benezra, 2003; Ying et al., 2003). These two factors can be thought of as balancing agents to the effects of Lif signaling. Dusp9 dephosphorylates Erk, inhibiting its pro-differentiation activity (Li et al., 2012). The Id genes play a role in blocking the predisposition of Lif cultured cells to differentiate toward

neuroectoderm lineages (Harland, 2000; Ying et al., 2003). Bmp4, on the other hand, drives the differentiation process toward mesodermal and endodermal lineages (Ying et al., 2003).

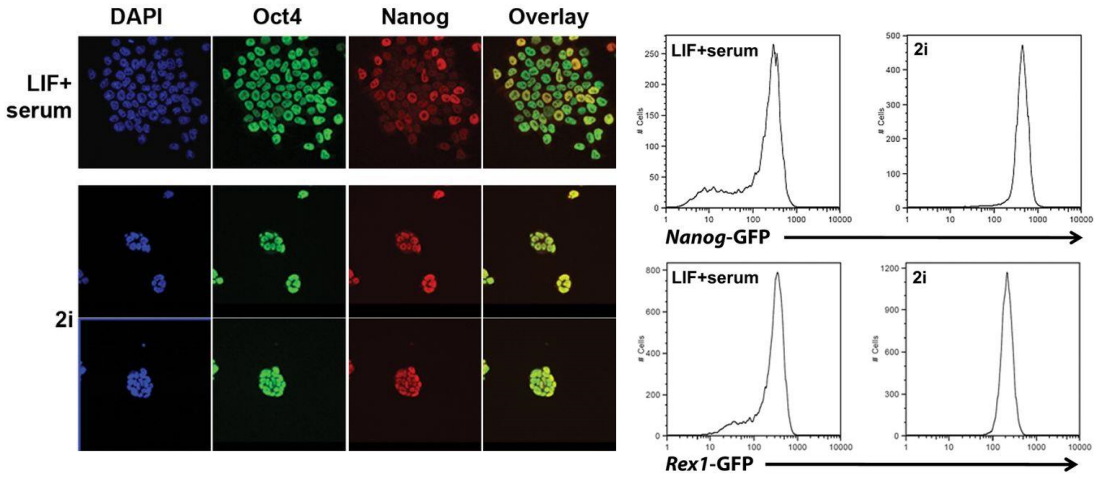


Fig. 1.5 Left, Immunostaining shows that the Lif+serum culture exhibits heterogeneity in key naïve markers such as Oct4 and Nanog, whereas 2i produces a more homogeneous starting population. Quantification by flow cytometry reveals a bimodal distribution of naïve markers Nanog and Rex1 in Lif+serum. The 2i system does not exhibit the lower valued mode and produces a cell population expressing these two markers with single modes in the higher range. (reproduced from Wray et al., 2010).

One feature of the serum/Lif culture system is the dynamic heterogeneity in expression levels of pluripotency factors (Fig. 1.5) (Chambers et al., 2007; Chickarmane et al., 2006; Glauche et al., 2010b; Kalmar et al., 2009; Toyooka et al., 2008). However, this heterogeneity is not observed in the pre-implantation epiblast, the *in vivo* analog of the ES cell (Boroviak et al., 2014, 2015). One of the confounding issues of serum/Lif is the unknown and variable composition of serum. Further work on culture systems focused on generating a homogeneous ES cell population, which more closely resembles the cells in epiblast of the pre-implantation embryo. This has been achieved by the two inhibitor (2i) system, a chemically defined basal medium supplemented with two small molecule inhibitors Chiron99021 (Chiron) and PD0325901 (PD03) (Wray et al., 2010; Ying et al., 2008). The 2i system has been shown to generate ES cell populations that closely resemble naïve marker expression levels found in the embryo and it reduces the heterogeneity of those expression levels observed in serum/LIF (Fig. 1.5) (Boroviak et al., 2014; Martello et al., 2013; Nichols and Smith, 2009, 2012; Wray et al., 2010; Ying et al., 2008).

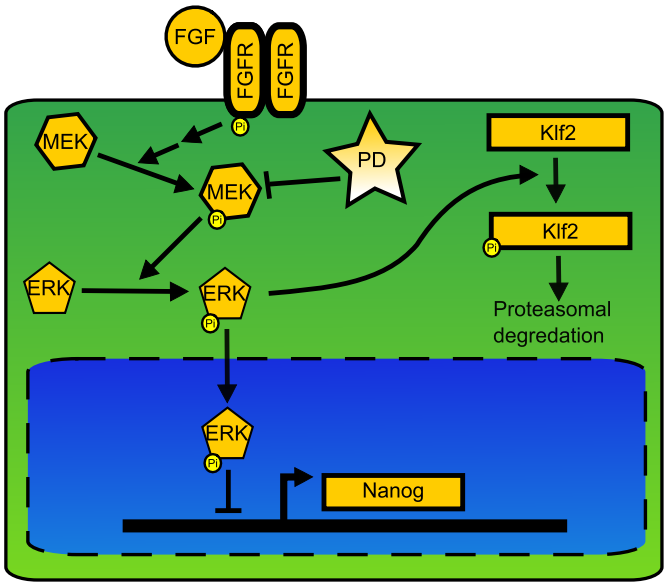


Fig. 1.6 Fgf-Mek/Erk signaling in ES cells. Fgf signaling is a major driver of differentiation. It functions by activation of Erk via the kinase cascade (Ras-Raf-Mek-Erk). Erk activation directly represses Nanog transcription and induces proteasomal degradation of Klf2 by phosphorylation. PD03 is a Mek inhibitor which is the only known upstream kinase of Erk. Inhibition of Mek effectively blocks the prodifferentiation activity of Erk. (adapted from Lodish, 2008; Quinlan, 2011).

PD03 is a Mek inhibitor. It effectively blocks fibroblast growth factor (Fgf) signaling from activating Erk (Fig. 1.6) (Ying et al., 2008). As described above, Erk promotes differentiation towards the neuroectoderm lineages. The Fgf signaling pathway is activated when Fgf4 binds to Fgf receptor (Fgfr), a receptor tyrosine kinase (RTK) (Lodish, 2008; Quinlan, 2011). Homodimerization stimulates the innate RTK activity of Fgfr which initiates the mitogen-activated protein kinase (MAPK) cascade. Two components of the MAPK cascade are Mek and Erk. Fgf4 had previously been identified as an inductive cue for differentiation *in vivo* and *in vitro* (Burdon et al., 1999b; Kunath et al., 2007; Lanner and Rossant, 2010). Blocking this inductive signal results in impaired differentiation (Burdon et al., 1999b; Kunath et al., 2007). Fgf signaling acts through phospho-Erk which suppresses Nanog transcription (Hamazaki et al., 2006; Yamanaka et al., 2010) and tags pluripotency factor Klf2 for proteasomal degradation (Yeo et al., 2014).

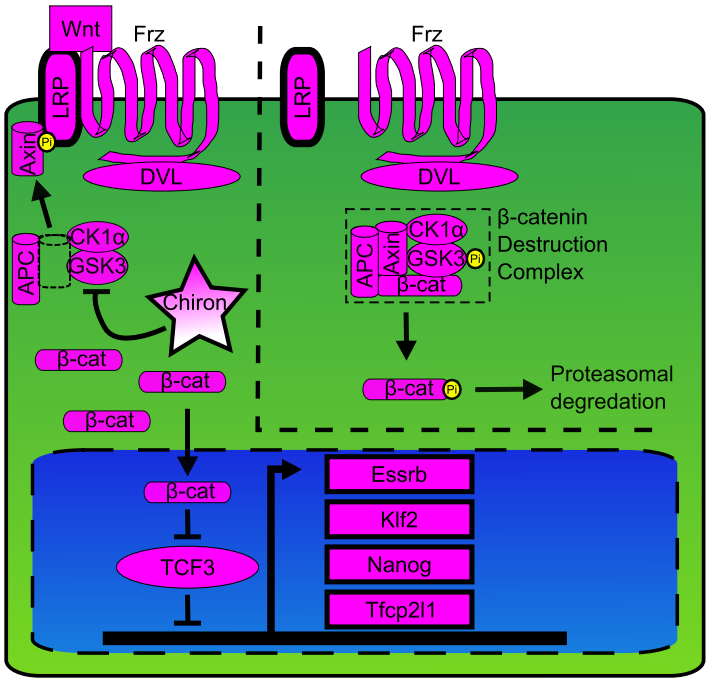


Fig. 1.7 Wnt-β-catenin signaling in ES cells. Tcf3 is a major transcriptional repressor of key naïve factors Essrb, Klf2, Nanog, and Tfcp2l1. Repression of Tcf3 by β-catenin allows for transcription of these factors. Tcf3 repression can be accomplished by inhibition of Gsk3 by Chiron or by Wnt signaling. In the absence of Chiron or Wnt, the β-catenin destruction complex phorphorylates β-catenin through Gsk3, which signals for its proetasomal degradation. The β-catenin destruction complex is disrupted in the presence of Wnt. The binding of Wnt to LRP and Frz results in the activation of LRP's RTK activity. This recruits Axin to the interior of the plasma membrane and the destabilization of the β-catenin destruction complex. Chiron, on the other hand, disrupts the β-catenin destruction complex by direct inhibition of Gsk3. Both mechanisms result in accumulation of β-catenin and inhibition of Tcf3. (adapted from Lodish, 2008; Quinlan, 2011).

Chiron is a glycogen synthase kinase (Gsk) 3 inhibitor and it is the second signal in the 2i system (Wray et al., 2010; Ying et al., 2008). Chiron can stimulate the canonical Wnt-signaling pathway by destabilizing the β-catenin destruction complex (Fig. 1.7). In the absence of Wnt, the subunits of the β-catenin destruction complex are associated. The subunits are: APC, Axin, CK1α, and Gsk3. The intact complex phosphorylates β-catenin, via Gsk3 (Doble et al., 2007), which signals for the polyubiquitination of β-catenin and its proteasomal degradation. On the other hand, Wnt signaling leads to the destabilization of this complex, beginning with the binding of Wnt to two cell surface receptors, the g-protein coupled receptors Frizzled and the RTK lipoprotein receptor-related protein (Lrp). Activation of the Lrp kinase activity causes dissociation of the β-catenin destruction complex by recruiting Axin to the cell surface. This allows free β-catenin to accumulate and translocate

to the nucleus where it directly binds the repressor Tcf7l1, also called Tcf3 (Pereira et al., 2006; Wray et al., 2011; Wu and Pan, 2010; Yi et al., 2008). This binding allows for the expression of core pluripotency factors such as Esrrb, Klf2, Nanog, and Tfcp2l1(Dunn et al., 2014; Martello et al., 2012; Yi et al., 2008). Chiron directly inhibits Gsk3, allowing for accumulation of β-catenin in the absence of Wnt (Wray et al., 2011; Yi et al., 2008).

Further studies showed that any combination of at least two of the signals Chiron, PD03, and Lif is sufficient to keep ES cells in this naïve state (Dunn et al., 2014; Wray et al., 2010). These combinations of signals also provide an environment that generates a more homogeneous cell culture, which narrows the distribution of pluripotency factor expression levels in individual cells (Wray et al., 2010; Ying et al., 2008). These extrinsic signals maintain ES cells in a pluripotent state by insulating the core pluripotency circuitry of transcription factors and chromatin modulators from insult (Dunn et al., 2014; Martello et al., 2013, 2012; Nichols and Smith, 2012). The actions of these extrinsic self-renewal signals are summarized in Fig. 1.8. Lif drives self-renewal by the activation of Jak/Stat3 pathway. Chiron is a Gsk3 inhibitor, a down-stream target of Wnt signaling. Inhibition of Gsk3 alleviates the suppressive power of Tcf3 via accumulation of β-catenin. PD03 inhibits Mek, which lies down stream of the FGF signaling pathway. Inhibition of MEK blocks the pro-differentiation activity of Erk.

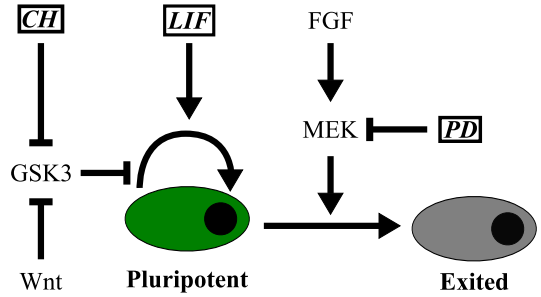


Fig. 1.8 The activity of Lif, Chiron (CH), and PD03 (PD). Lif is a trophic factor that promotes self-renewal by stimulation of the Jak-Stat3 pathway, this in turn induces the transcription of naïve factors. Chiron is a Gsk3 inhibitor that stops the proteasomal degredation of β-catenin, which then alleviates the transcriptional repression of naïve factors by Tcf3. PD03 is a Mek inhibitor the effectively block the prodifferentiation activity of Erk. Erk directly block the transcription of naïve factors and labels others for proteasomal degradation by phosphorylation. Any combination of at least two of these signals is sufficient to maintain ES cell self-renewal. For this work, the 2i system is used, Chiron + PD03.

1.2 Exit from Naïve Pluripotency

The ability to establish and maintain pluripotency *in vitro* holds potential for disease modeling, drug screening, and cell based therapies (Smith, 1998). However this potential can only be realized with effective *in vitro* differentiation. Differentiation is the irreversible process in which a stem cell loses the ability to self-renew and acquires a new, more restricted identity with novel properties (Martello and Smith, 2014). A better understanding of the developmental signals responsible for embryo patterning will aide in effectively controlling *in vitro* differentiation (Smith, 1998).

In the embryo, naive pluripotency is first acquired in the ICM of the blastocyst upon epiblast specification and is maintained in the epiblast of the E4.5 late blastocyst stage until implantation (Fig. 1.9) (Boroviak and Nichols, 2014; Cockburn and Rossant, 2010). The early ICM co-expresses markers for both the epiblast and the primitive endoderm, the source of the yolk sack (Boroviak et al., 2014; Guo et al., 2010; Ohnishi et al., 2013). Primitive endoderm specification is induced by Fgf4, which is secreted by cells acquiring the epiblast identity (Nichols and Smith, 2009; Yamanaka et al., 2010). It is thought that extracellular matrix proteins produced by the emerging primitive endoderm reinforce the nascent epiblast identity by up-regulation of Nanog and Sox2 (Boroviak et al., 2014). Specification of the early ICM into epiblast and primitive endoderm occurs asynchronously, where the pool of common progenitor cells is incrementally allocated to one of the cell fates (Saiz et al., 2016). Asynchronous down-regulation of Nanog is also observed *in vivo* peri-implantation between the E4.5 and E4.7 stages (Acampora et al., 2016). However, in both instances, these data are unable to discriminate the origins of the asynchronicity. Asynchronus specification can occur by asymmetric divisions, where one stem cell divides into one stem cell and one differentiated cell. Alternatively, it can also arise from asynchronous symmetric divisions, where a symmetric division occurs followed by differentiation at different times.

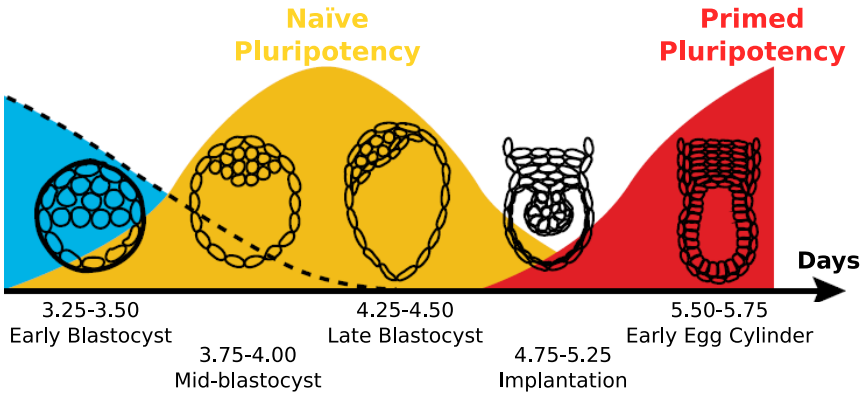


Fig. 1.9 Naïve pluripotency emerges in the epiblast of the the E3.75 mid-blastocyst. The inner cell mass (ICM) of the E3.25 Early blastocyst co-expresses markers for the epiblast and the primitive endoderm. Secretion of laminin by nascent primitive endoderm cells induces ICM cells to acquire the Epiblast identity by down-regulating primitive endoderm markers and up-regulating the naïve factors. This identity is maintained through the E4.25 late blastocyst stage. Upon implantation, the naïve circuitry is shut off while the primed pluripotency circuitry is activated. (adapted from Boroviak and Nichols, 2014).

Pre-implantation epiblast and ES cells exhibit a high similarity in their transcriptional profiles (Boroviak et al., 2014). In normal development, progression away from naïve pluripotency towards a primed state starts at implantation (Fig. 1.9) (Boroviak and Nichols, 2014). Upon implantation, naïve factors are down-regulated (Boroviak et al., 2014) and markers for lineage priming are up-regulated (Boroviak and Nichols, 2014; Kalkan et al., 2017; Mulas et al., 2017). Secretions from the extra-embryonic tissues, including Bmp and Fgf, are driving factors for the exit from naïve pluripotency and beyond (Arman et al., 1998; Boroviak et al., 2015; ten Berge et al., 2011; Ying et al., 2003).

In ES cells, the transition to exit from the naïve state begins with the termination of self-renewal and the initiation of commitment (Betschinger, 2017; Martello and Smith, 2014). Commitment is associated with the dissolution of the pluripotency transcription factor network (Fig. 1.10). This requires a coordinated down-regulation of naïve factors and up-regulation of specific primed-pluripotency factors (Betschinger, 2017; Martello and Smith, 2014). Once exit has occurred, cells are incapable of autonomous reversion to the previous, less restricted identity (Kalkan et al., 2017). However, this reversion can be induced by reprogramming, a subject beyond the scope of this work (Okita et al., 2007; Takahashi et al., 2007; Takahashi and Yamanaka, 2006).

Exit in ES cells begins upon withdrawal of self-renewal signals. At the point of exit, the molecular network that sustains naïve pluripotency is dismantled (Buecker et al., 2014; Davies et al., 2013; Kalkan and Smith, 2014; Leeb et al., 2014). Lineage specification of ES

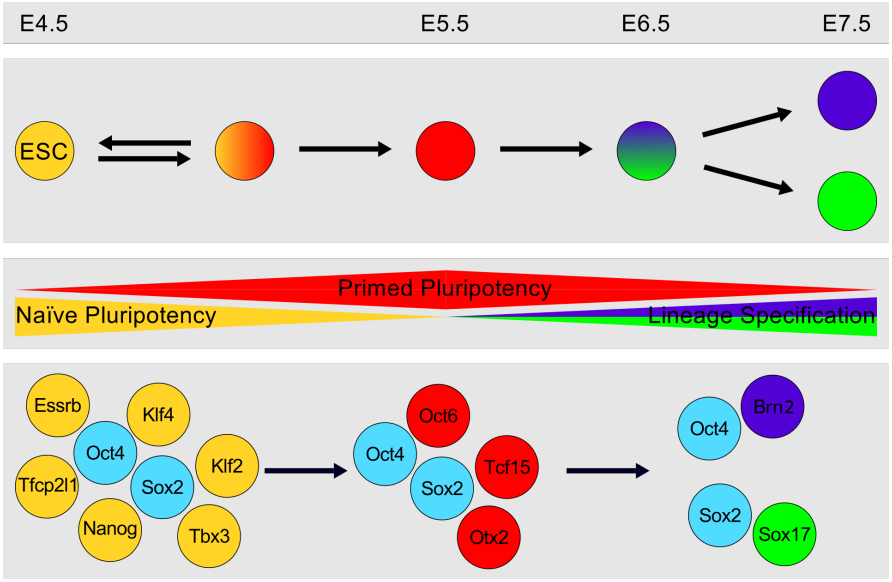


Fig. 1.10 The core pluripotency factors Oct4 and Sox2 sit at the heart of the pluripotency transcription factor network (blue). ES cells transition out of naïve pluripotency by the controlled down-regulation of naïve factors (yellow) and the up-regulation of post-implantation, or primed, factors (red). This must occur before cells are able to up-regulate lineage specific markers like Brn2 (purple) in neural stem cells and Sox17 (green) in endodermal cells. (adapted from Betschinger, 2017, and Martello and Smith, 2014).

cells in serum/Lif is suggested to proceed in a probabilistic fashion (Moris et al., 2016). The culture yields a dynamic heterogeneity where cells reversibly lose the ability of self-renewal, with simultaneous down-regulation of naïve factors and up-regulation of lineage specific markers (Toyooka et al., 2008). It is thought that this process allows ES cells to explore a space of possible cell fates, priming them for lineage specification (Betschinger, 2017; Moris et al., 2016; Torres-Padilla and Chambers, 2014; Toyooka et al., 2008). However, this heterogeneity has been attributed to the culture system rather than an innate trait of development (Marks et al., 2012). ES cells in the 2i system do not exhibit this behavior and are still capable of contributing to germline competent chimeras without this transient, reversible priming (Nichols and Smith, 2009).

It has been established that ES cell culture systems ensure self-renewal by isolating the core pluripotency circuitry from destabilizing factors (Nichols and Smith, 2012). Withdrawal of 2i allows for the reactivation of pro-differentiation Fgf signaling (Burdon et al., 1999b; Ying et al., 2008) and the repressive activity of Tcf3, down stream of β-catenin (Pereira et al., 2006; Wray et al., 2011; Wu and Pan, 2010; Yi et al., 2008). While, elimination of key factors, like Esrrb or Nanog, destabilizes the transcription factor network, making ES cells

prone to spontaneous differentiation (Chambers et al., 2007; Martello et al., 2012).

1.2.1 Mechanisms in the Dissolution of the Naïve Pluripotency Transcription Factor Network

Exit from naïve pluripotency upon withdrawal of extrinsic self-renewal signals is not simply the passive down-regulation of naïve factors. Rather the dissolution of naïve pluripotency transcription factor network occurs in a well coordinated manner using a battery of repressive regulators. In fact, many of these negative regulators are present in ES cells, placing ES cells in a position poised for differentiation (Betschinger, 2017; Kalkan and Smith, 2014). These regulators work in concert to silence the naïve factors while allowing for the activation of the primed network (Kalkan and Smith, 2014).

One of the major negative regulators of the naïve pluripotency transcription factor network is Tcf3 (Guo et al., 2011; Yi et al., 2008). Tcf3 is a direct repressor of many key naïve pluripotency transcription factors including Essrb, Nanog, Tfcp2l1, and Klf2 (Martello et al., 2012; Pereira et al., 2006). As described in Sec. 1.1 (Fig 1.7), inhibition of Gsk3 alleviates the repressive activity of Tcf3. However, Tcf3 activity is able to be overridden by enforced nuclear localization of Tfe3. This maintains pluripotency in the absence of exogenous signals through expression of Esrrb. Upon withdrawal of self-renewal signals in WT ES cells, Flcn-fnip1/2 and Tsc1/2 converge to relocate Tfe3 out of the nucleus, resulting in the down-regulation of Esrrb (Betschinger et al., 2013).

A study using haploid ES cell has shown that negative regulation can occur at the transcriptional and post-transcription levels. Zfp706 blocks the transcription of Klf4 (Leeb et al., 2014). When Zfp706 is deleted, ES cells display and inhibited capacity to differentiate. Upon deletion of Klf4, Zfp706-null cells recover their differentiation capacity (Betschinger et al., 2013). In that same study, the RNA binding protein Pum1 was identified as post-transcriptional repressor of naïve pluripotency. Pum1 destabilizes the naïve pluripotency circuitry by binding the 3'-UTR of Tbx3, Tfcp2l1, Klf2 and Esrrb transcripts which inhibits their translation (Leeb et al., 2014).

Global changes in the epigenome are seen upon transition out of the naïve state (Betschinger et al., 2013; Kalkan and Smith, 2014). During early transition Dnmt3a1 and Dnmt3b are up-regulated and necessary for the methylation changes observed (Martello and Smith, 2014). Addition chromatin modifiers negatively regulate the naïve network. The nuclesome remodeling and histone deacetylation co-repressor complex NuRD is needed for ES cell

differentiation (Kaji et al., 2006, 2007). Deletion of the NuRD scaffold protein Mbd3 results in the up-regulation of naïve factors Tbx3 and Klf4 (Reynolds et al., 2012). NuRd also plays a role in Nanog regulation where it is recruited to the Nanog promoter by Zfp281 (Betschinger et al., 2013).

Fgf/Erk signalling is known to be a major driver of exit from naïve naive pluripotency (Yamanaka et al., 2010; Ying et al., 2008). Erk is able to directly effect naïve transcrition factors, such as Klf2 and Nanog (see. Fig 1.6). A recent study has shown that Erk activity is also responsible for the nuclear export of naïve factor Klf4 (Dhaliwal et al., 2018). Fgf signaling also activates the priming activity of Tcf15, which is linked to the down-regulation of Nanog and the up-regulation of primed fator Otx2 (Davies et al., 2013). Oct4 and Otx2 are through to be responsible for bringing the primed network online through co-operative binding (Kalkan and Smith, 2014).

The naïve pluripotency transcription factor form a highly redundant and self-stabilizing newtwork. The timely dissolution of this network requires a cadre of negative regulators working in unison. However, the exact mechanism(s) by which the naïve network is take offline is still an area for discovery.

1.2.2 The Role of the Cell-Cycle in Differentiation

In hPSCs, the cell-cycle is implicated in lineage specification. The control of endoderm and neuroectoderm cell fate decisions are mediated through Smad signaling (Pauklin and Vallier, 2013, 2014). During the early G1 phase, the absence of Cyclin-D allows for Smad2/3 nuclear localization. Smad2/3 directly binds to and activates endoderm genes (Pauklin and Vallier, 2013, 2014). In the late G1 and the G1/S phase, as defined by the FUCCI system, Cyclin-D is present. Activation of CDK3/4 results in the phosphorylation and nuclear exclusion of Smad2/3. This prevents induction of endoderm differentiation and is permissive of neuroectodermal specification (Pauklin and Vallier, 2013, 2014).

The cell-cycle has also been suggested to be an important regulatory step in ES cell differentiation as well. Both the inhibition of cell division and DNA replicaiton delay exit from naïve pluripotency (Waisman et al., 2017). Inihibiton of cell division resulted in slower kinetics of up-regulation in primed markers Fgf5, Oct6, and Dnmt3a and down-regulation of naïve marker Nanog (Waisman et al., 2017). While Cells with blocked DNA replication saw additional retardation of up-regulation in primed marker Otx2 and down-regulation in naïve markers Klf4, and Essrb. When cells were allowed to resume DNA replication, a moderate

return to the WT expression kinetics was observed. Thus, completing rounds of cell-cycling is not necessary for exit, however it enhances the progression out of the naïve state (Waisman et al., 2017).

During development, a progressive acceleration in the cell cycle is seen. Snow, 1977, notes that at the onset of the primitive streak formation, the embryo enters "a period of rapid growth". For about 24 hours, small-regions of the epiblast (10%), termed "proliferative zones", exhibit accelerated cycling with cell-cycles shorter than 4 hours (Snow, 1977). These highly proliferative areas are responsible for generating the majority of the tissue in the E7.5 epiblast. This rate of cycling is more rapid than earlier stages of development, including the E4.5 which has a generation time of about 9 hours.

In ES cells, it has been shown that inhibition of Fgf signaling results in a prolonged G1 phase (ter Huurne et al., 2017). This is thought to be a direct result of Erk inhibition by PD03. Erk is an activator of the CDK/Cyclins and is required for the entry into the cell cycle (Chambard et al., 2007). Normally, the CDKs would act to block the repressor RB, to allow E2f transcription and passage through the G1 restriction point. This begs the question, is there a similar progressive shortening of the cell-cycle upon withdrawal of 2i, much like that of the embryo during progression through the stages of development?

1.2.3 Divisions Symmetry During Differentiation

At each round of division, stem cells will either self-renew or differentiate. The symmetry of division is defined by the fates of the daughter cells (Morrison and Kimble, 2006). In a symmetric division, stem cells will produce two equivalent daughter cells that either retain their stem cell identity or take on a new more restricted identity with novel attributes. In an asymmetric division, the two daughter cells assume different identities. Asymmetric divisions are controlled by either extrinsic or intrinsic regulation (Betschinger and Knoblich, 2004; Morrison and Kimble, 2006). Extrinsic regulation of asymmetric divisions occurs when the two daughter cells encounter two different environments, whereas intrinsic regulation occurs when cell fate determinants are differentially inherited (Betschinger and Knoblich, 2004; Morrison and Kimble, 2006).

Extrinsically regulated asymmetric divisions occur when two daughter cells are born with the same potential, but then encounter different environments(Betschinger and Knoblich, 2004; Morrison and Kimble, 2006). The difference in environmental cues will cause the daughter cells to have different cell identities. In adult tissues this happens when a stem cell migrates from the niche. This cell no longer receives the signals required to self-renewal and

with inevitably differentiate. An example of this can be found in the Drosophila germline stem cells (GSCs) (Yamashita et al., 2005). In the ovary, the niche is comprised of cap cells which produce the Bmp2/4 homolog decapentaplegic (Dpp) which stimulates the Tgfβ pathway in female GSC. Tgfβ in turn represses the pro-differentiation gene bag-of-marbles (Bam). In Drosphilia testes, hub cells form the niche. Hub cells produces Unpaired (Upd) which activates pro-self-renewal Jak-Stat signalling in male GSC. In both instances, once a GSC migrates from the niche, it will stop receiving the inductive self-rewnal signals, and will inevitably differentiate into either female cyctoblasts or male gonialblasts (Yamashita et al., 2005).

Two features are required for this type of asymmetric division (Yamashita et al., 2005). First, the signals from the niche must regulate self-renewal. Second, the stem cells must remain in close proximity to the niche for the maintenance of their identity. In fact, it has been hypothesized that adherens junctions are important in maintaining the proximity of the stem cells to the niche. This is demonstrated in the concentration of E-cadherin homologue Shotgun and β-catenin homologue Armadillo at the interface of GSCs and either cap cells in females or hub cells in males (Song et al., 2002; Yamashita et al., 2003).

Intrinsically regulated asymmetric cell divisions occur at the birth as a result of asymmetric allocation of cell fate determinant in the two daughter cells. Betschinger and Knoblich, 2004, explain that three features must be in place for intrinsically regulated asymmetric cell divisions to occur. First, an axis of polarity must be generated within the mother cell, second, the mitotic spindle must be oriented parallel to the axis of symmetry, and third, cell fate determinants must be unequally distributed along this axis at the time of cell division. Two examples of intrinsically regulated asymmetric division are the C. *elegans* embryo and Drosophila neuroblast. Both of the these cell types employ Par3/6 as to establish and maintain cellular polarity (Betschinger and Knoblich, 2004; Clevers, 2005; Doe and Bowerman, 2001; Gönczy and Rose, 2005; Rose and Gönczy, 2014).

In the C. *elegans* embryo, the axis of polarity is established upon fertilization of the egg by the sperm, which induces anterior localization of Par-3/6 (Gönczy and Rose, 2005; Rose and Gönczy, 2014). Cellular polarity is maintained through mutual inhibition of Par-3/6 and Par-2, which is localized in the posterior (Gönczy and Rose, 2005; Rose and Gönczy, 2014). Cell fate determinants are asymmetrically segregated by cytoplasmic flow induced by protein degradation and diffusion. Kinase Par-1 excludes Mex-5/6 from posterior cytoplasm by phosphorylation and subsequent proteosomal degradation. Anterior Mex-5/6 then diffuse to the posterior, inducing an anterior-to-posterior cytoplasmic flow that carries cell fate determinants Pie-1, Pos-1, Mex-1, and p-granules to the posterior (Betschinger and Knoblich,

2004). Anterior/posterior cell divisions then results in the anterior AB-cell, ectodermal precursor, and the posterior P1 cell, which will give rise to the endoderm, mesoderm, and germ line (Doe and Bowerman, 2001).

Drosophila neuroblast are derived from internally delaminating neuroepithelium. This Neuroepithelium has an inherent apicobasal polarity. Par-3/6 and aPKC are located on the apical domain and, upon delamination, the apical domain remains in contact with the basal neuroepithelium. Apical polarization is maintained by Inscuteable, and its adapter partner of inscuteable, which induce neuroblast specification. (Morrison and Kimble, 2006). Basal polarization is maintained by Prospero, and its adapter protein Miranda, and Numb, and its adapter protein partner of numb. The asymmetric inheritance of numb induces down regulation of notch signaling resulting in ganglion mother cell specification (Betschinger and Knoblich, 2004; Clevers, 2005). Indeed, this is an intrinsically regulated asymmetric cell divisions because neuroblast cells will continue to divide asymmetrically as single cells (Clevers, 2005).

Examples of intrinsic asymmetric division can be found in several mammalian systems including myogenesis, neurogenesis, and epithelial stratification. The regulation of Notch signaling is a mechanism that is conserved in both myogenesis and neurogenesis. Asymmetric segregation of Numb in satellite cells (Conboy and Rando, 2002; Tajbakhsh et al., 2009) and neural precursors (Zhong et al., 1997) result in differentiation. Furthermore, the regulation of the mitotic spindle in neural precursors (Chenn and McConnell, 1995) and basal epithelium are needed (Lechler and Fuchs, 2005). Proliferative basal epithelial cells require asymmetric allocation of apical α-catenin for the recruitment of aPKC, Par3–LGN–Inscuteable complex, and NuMA–dynactin for spindle alignment. Division along the axis results in one committed cell that will stop dividing and stratify apically and one basal cell that will continue to proliferate (Lechler and Fuchs, 2005).

They symmetry of cell divisions can determined by extrinsic and intrinsic factors. Extrinsically regulated asymmetric cell divisions require a structure, like the niche, to creates an inhomogeneous environment, where asymmetry is induced by migration of a daughter cell. Intrinsically regulated asymmetric cell divisions require the establishment and maintenance of polarization through the end of the cell-cycle. Thus, in the case of *in vitro* differentiation of ES cells, non-polarized cells in a seemingly homogeneous environment, what type of cell divisions do they employ?

1.3 Long-Term Single-Cell Imaging and Tracking

Most of what is know about this transition to exit comes by way of population averages (macromolecular blotting, PCR, microarray) or single-cell snapshot data (clonal imaging, flow cytometry, single-cell -omics) (Schroeder, 2011). Population averaging causes a loss of information regarding the state of individual cells (Fig. 1.11). This is satisfactory for studying homogeneous cell populations, but becomes an issue where multimodal populations arise. Single-cell snapshot (Fig. 1.11) data improves upon population averages with the ability to reveal these subpopulations. However, these types of data do not preserve genealogical or kinetic information and therefore cannot elucidate how a heterogeneous population has come into being. Long-term single-cell tracking (Fig. 1.11) preserves this information, which confers the ability to discriminate if this is occurring within a single clone or between clones.

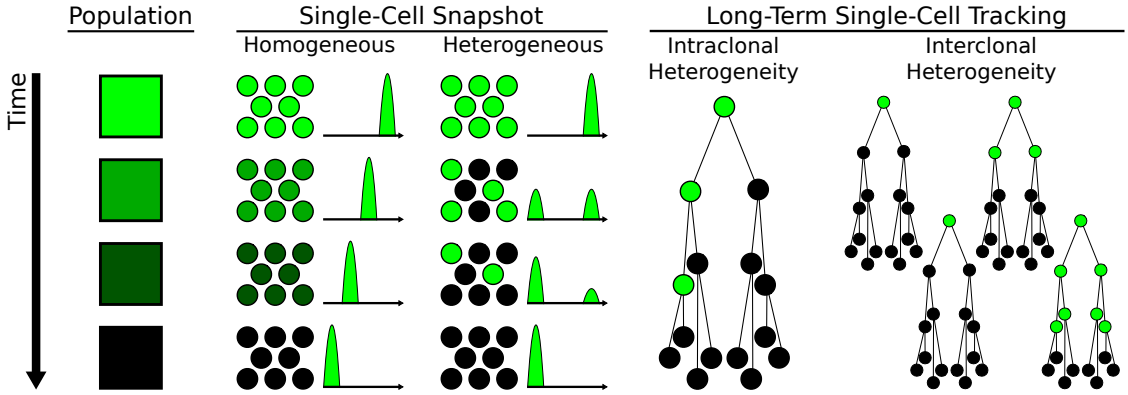


Fig. 1.11 Long-term single-cell imaging and tracking provides a greater resolution of cellular information than other methods. Population averaging techniques, like macromolecular blotting, generate data that have the ability to show general trends. However, these methods destroy information about individual cells and are thus unable to discriminate cellular subpopulations. Single-cell snapshot techniques, such as high-resolution imaging of fixed cells or single cell -omics, retain information about individual cells, but destroy the temporal information. Particularly the history and future of individual cells, and thus these data lack the ability to show how subpopulations arise. Long-term single-cell tracking preserves both single-cell and temporal information. (adapted from Schroeder, 2011).

The core experiment of this thesis is the observation of ES cells as they down-regulate the Rex1-GFPd2 reporter during the transition to exit from naïve pluripotency (Fig. 1.12 A). The extensive use of time-destructive methods have left much of the cellular dynamics of exit unexplored. Long-term single-cell imaging and tracking over many generations will begin to resolve some of the remaining questions. Kalkan et al., 2017, have noted the exit process to be asynchronous, which implies a level of heterogeneity in this cell

population. However, 2i ES cells appear to be homogeneous in their expression of naïve factors. Long-term single-cell imaging and tracking data can shed light on how a seemingly homogeneous population become heterogeneous. It offers the ability to measure the kinetics of Rex1-GFPd2 down-regulation in near real time (Fig. 1.12 B). These data can then be used to reconstruct lineages to determine the origin of the asynchronicity of exit by examining cell fates and the symmetry of divisions (Fig. 1.12 C). Genealogical information can be used to ask if this heterogeneity manifests itself in other cellular attributes like the cell-cycle, motility, and morphology (Fig. 1.12 D).

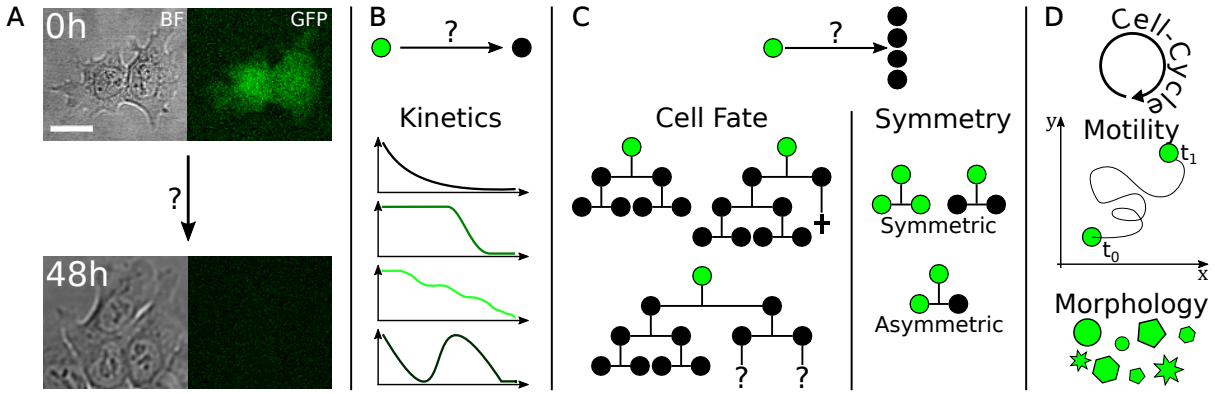


Fig. 1.12 (A) Long-term single-cell imaging was used to directly observe cellular kinetics associated exit from naïve pluripotency. (B) Near real-time quantification of the Rex1-GFPd2 will allow for the quantification of reporter down-regulation in exiting ES cells. (C) Lineage-trees can be constructed (C, left), showing when cells exit, as well as the symmetry of division (C, right). (D) Kinetic information regarding cell behaviors like the cell-cycle, motility, and morphology can be quantified (adapted from Etzrodt et al., 2014)

Long-term single-cell tracking is a highly interdisciplinary method which demands expertise from molecular biology, microscopy, computer science, mathematics, and data science (Skylaki et al., 2016). The pipeline for long-term single-cell imaging begins with the cell (Fig. 1.13). Cells must be able to be tracked over many generations, which are often in dense clusters, and the construction of cell lines with a tracking reporter makes the process of raw image analysis more efficient. This is accomplished by labeling subcellular structures, like the nucleus or plasma membrane, with constitutively expressed fluorescent proteins. Using the reporter of interest, in this case Rex1-GFPd2, as a substitute for the tracking reporter is problematic because it eventually shuts off. The tracking reporter must create clear cellular boundaries and maintain fluorescence levels for the duration of imaging. It is also important to construct cell lines for controls, which are often not used. The most likely reasons are the prohibitive cost of imaging and the time intensity of tracking. Rather, synthetic negative controls have been constructed based upon the image background

(Filipczyk et al., 2015). This may be acceptable for negative controls (signal "off"), obut this is not sufficient to determine when a reporter is "on"; this requires a true positive control.

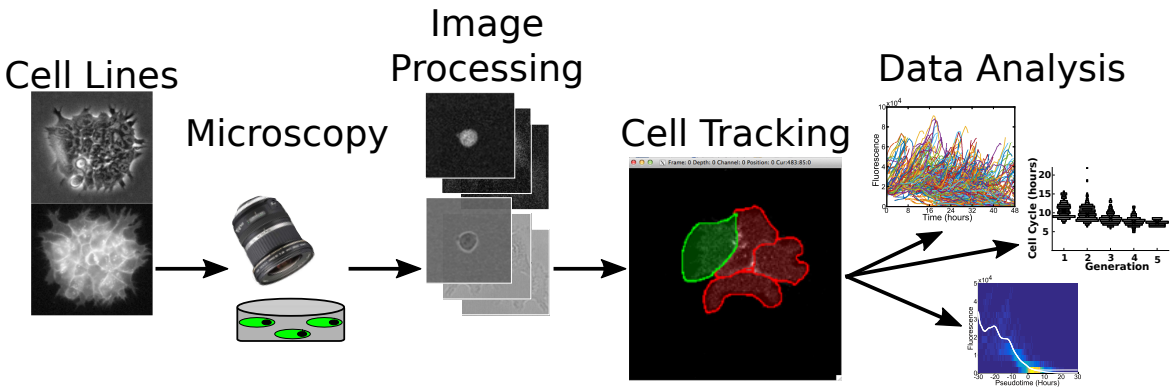


Fig. 1.13 The long-term single-cell imaging pipeline. Appropriate cell lines must be constructed for efficient tracking and experimental controls. The microscopy environment must be able to maintain a healthy cell culture so that stressed cells do not result in bad data. Imaging files must be stored and processed into formats which can be used for single-cell tracking. Various single-cell tracking platforms exist, however the most reliable are currently human supervised semi-automated software. Once the tracking data has been collected, it must be analyzed; this often require constructing custom software for the specific analyzes being performed.

Once appropriate cell lines have been constructed, cells need to be imaged for the duration of the process being observed. I am observing the down-regulation of a pluripotency marker, therefore imaging must occur over a long enough period of time that silencing is observed in the majority of cells. These cultures will potentially be imaged for many days and will require the basic needs of tissue culture. Construction of a heated and humidified incubator microscope is essential. Specification for these can be found in several reviews (Coutu and Schroeder, 2013; Muzzey and van Oudenaarden, 2009; Schroeder, 2011; Skylaki et al., 2016) and commercially built incubator microscopes are available for purchase. However, the latter option reduces the flexibility of modification for different assays. During imaging, a balance must be struck between the quality of images and the health of the culture. As Schroeder, 2011, points out, it is better to have bad images of healthy cells than excellent images of dead cells. This balance is maintained by reduction in imaging frequency, use of minimal laser intensity, and fluorophore selection. The frequency of imaging must be on the same time scale as that of cell motility, laser intensity should be enough to maintain a sufficient signal to noise ratio, and fluorophores with low energy excitation are ideal. Once imaging is complete, the data must be processed, stored, and converted into a format which can then be used for single-cell tracking.

The main bottle neck that still exists in this technique is the tracking stage (Fig. 1.13). Currently there are several options to perform single-cell tracking, with ranging levels of automation. A fully manual method may appear to be the fastest and easiest method for small projects, requiring a simple image processing tool such as ImageJ. One must go through each image separately, manually segment each cell, and record the kinetic and genealogical information by hand or in a text file. This methods is time consuming and thus hard to scale to large data sets.

Semi-automated software is currently the best means of cell tracking. The Tracking Tool (TTT), formerly Timm's Tracking Tool (Hilsenbeck et al., 2016), and bespoke tracking and analysis software from Blanchard et al., 2009, are suitable examples. Semi-automation refers to the fact that the user is receiving some form of computer assistance. The automation is the collection and recording of kinetic and genealogical data. These software also have the added benefit of modifiable cell masks which highlight cells that have already been traced. This prevents the user from mistaking cell identities and from double counting pixels by showing the user what part of an image has been allocated to which cell.

Fully automated segmentation and genealogy recording has been accomplished for cells with regular morphologies like epithelium (Blanchard et al., 2009; Kursawe et al., 2016) and the modern version of TTT has the ability to interface with automated segmentation software, like CellProfiler. Modern segmentation algorithms are sophisticated enough to identify an individual person from satellite images, yet a universal fully automated and unsupervised cell tracking software is still lacking. The issue comes from the balancing of image quality and frequency of imaging with health of the cell culture. So, until imaging techniques are able to produce adequate images at a high enough time resolution without damaging cells, there is still need for human supervision.

Once the tracking has been completed, these data now must be analyzed. Another limitation in long-term single-cell tracking is the data analysis (Skylaki et al., 2016). Handling and analysis of the complex data are nontrivial. Each batch of imaging data must be analyzed using purpose built libraries of functions, limiting their use across different systems (Ulman et al., 2017; Winter et al., 2011). The work presented in this thesis required modifying existing tracking software (Blanchard et al., 2009) and constructing a library of MATLAB functions to analyze the resulting tracking data.

1.3.1 Rex1 (ZFP42) as a Marker for Pluripotency

For this thesis, the pluripotency marker Rex1, also called Zfp42, is used to monitor the transition to exit. This marker has been used in several studies to examine pluripotency (Herberg et al., 2016; Nakai-Futatsugi and Niwa, 2016; Wray et al., 2010) and the exit from naïve pluripotency (Kalkan et al., 2017; Kalkan and Smith, 2014; Mulas et al., 2017). Specifically, the Rex1-destabilized GFP (Rex1-GFPd2) cell line is used.

This has been generated by replacing the coding portion of one Rex1 allele with the sequence coding for GFPd2, a destabilized form of GFP (half-life $\approx$ 2 hours). The GFP is destabilized with the addition of two PEST sequences that increase its rate of proteasomal degradation (Kalkan et al., 2017; Wray et al., 2011). No disruption is seen from this knock-in, as Rex1 is dispensable in the early embryo and shows no ill effect on ES cells via knock-out or over expression (Kim et al., 2011; Masui et al., 2008; Rezende et al., 2011).

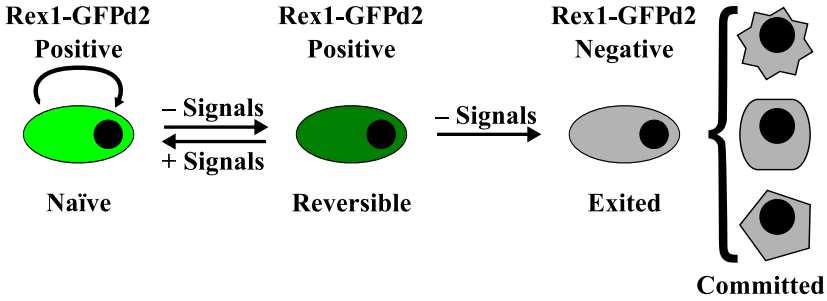


Fig. 1.14 ES cells will remain in the pluripotent state so long as the core pluripotency machinery in stimulated by select combinations of exogenous signals. When these signals are withdrawn, cells down regulate the naïve pluripotency circuitry. This can be monitored in near real time by Rex1-GFPd2. While the cells are reversible, signals can be reapplied and cells can be recovered back to the naïve state. Once cells have exited, reapplication of signals has no effect, as the cells have already exited.

Exit from the naïve pluripotent state involves shutting down the pluripotency circuitry (Dunn et al., 2014; Kalkan and Smith, 2014; Masui et al., 2008; Toyooka et al., 2008). *In vitro*, this process begins upon withdrawal of self-renewal signals. The status of the ES cell can be monitored via the Rex1-GFPd2 reporter in near real-time (Fig. 1.14) (Kalkan et al., 2017; Mulas et al., 2017). Upon initial signal withdrawal, there exists a window of time in which these cells dwell in a reversible state. These reversible cells can be brought back to naïve pluripotency with the reapplication of self-renewal signals. However, once the cells have shut off the naïve pluripotency circuitry, simple reapplication of these signals is not sufficient for recovery (Kalkan et al., 2017). It is at this point that ES cells have exited naïve pluripotency.

Rex1-GFPd2 has been demonstrated to be a highly sensitive reporter that marks naïve pluripotency (Kalkan et al., 2017; Kalkan and Smith, 2014; Mulas et al., 2017). Rex1 is a target of many pluripotency factors, including Nanog, and it is tightly associated with their down-regulation (Chen et al., 2008; Kalkan et al., 2017; Kim et al., 2008; Shi et al., 2006; Toyooka et al., 2008). Furthermore, this system faithfully reports endogenous Rex1 in near real-time (Boroviak et al., 2015; Boroviak and Nichols, 2014; Kalkan et al., 2017; Pelton et al., 2002). The combination of the two hour half-life of GFPd2 and the tight association with the naïve circuitry make this system ideal for long-term single-cell imaging.

1.4 Dissertation Aims

How cell fate transitions occur is a fundamental question in developmental biology. Much of what is known about the transition to exit from naïve pluripotency comes from population data and snap-shots of single-cell data. These have provided a wealth of information with regard to the molecular machinery underlying the maintenance of the pluripotency transcription factor network and, to a lesser extent, the dissolution of this network upon exit. However, this phenomena has not been been continuously observed on the single-cell level (Martello and Smith, 2014). My main aim is to study the dynamics and the factors affecting the irreversible exit from naïve pluripotency. This includes cell size and motility, the relationship to the cell-cycle, the symmetry of cell divisions, and the real-time kinetics and synchronicity of Rex1-GFPd2 down-regulation at single cell resolution.

Aims:

- How do cell-cycle duration, motility, and cell size change during exit?
- How do single cells down-regulate Rex1-GFPd2?
- How do ES cells divide during the exit from naïve pluripotency: symmetric, asymmetric, or mixed?
- Does the asynchronicity of exit arise within individual colonies or only within the population?

How do cell-cycle duration, motility, and cell size change during exit?

2i ES cells exist in a type of steady-state, poised for differentiation into different cell lineages. Static images of ES cell colonies show them to be tightly joined units (Andersson-Rolf et al., 2017; Wray et al., 2010). A different image appears when this steady-state is perturbed by the withdrawal of the signals which maintain self-renewal. Differentiated colonies appear to be large, loosely associated aggregates (Andersson-Rolf et al., 2017; Wray et al., 2010).

The exit from naïve pluripotency is associated with the well regulated dissolution of the naïve circuitry and the activation and establishment of the primed pluripotency circuitry (Betschinger, 2017; Kalkan et al., 2017). In the 2i system, this transition begins upon withdrawal of the two inhibitors, which allows for the reactivation of Mek/Erk signaling and the re-association of the β-catenin destruction complex.

Erk signaling plays a role in a wide array of cellular activities including cell-cycling and motility (Chambard et al., 2007; Tanimura and Takeda, 2017). Erk activity is required to allow the cell to pass the G1/S checkpoint (Chambard et al., 2007). In ES cells, there is evidence to suggest that inhibition of Mek, the upstream effector of Erk, by PD03 causes a prolonged G1 phase (ter Huurne et al., 2017). So is it the case that upon withdrawal of PD03, the reactivation of Erk activity results in faster cell-cycling?

ERK signaling also stimulates cell motility by direct phosphorylation of proteins involved in the formation of focal adhesion (cell-ECM adhesions) and locomotion (Tanimura and Takeda, 2017). During exit, there are changes in the expression levels of mRNAs that code for proteins involved in motility and for secreted ECM proteins (Kalkan et al., 2017). This includes, but is not limited to, down-regulation in expression of mRNAs coding for Myh14, Fn1, Lama1, and Lamc1 and up-regulation in Myh10 and Col18a (Kalkan et al., 2017). The function of these genes in the context of motility in ES cells and during exit is not immediately clear, but, given the indication of differential expression of genes coding for locomotion and for ECM proteins, pre- and post- withdrawal of self-renewal factors, coupled with the known stimulatory effect of Erk signaling, suggests that cell motility may increase during the exit process?

During self-renewal in 2i, Chiron inhibits GSK3 leading to the dissociation of the β-catenin destruction complex. This is followed by a subsequent increase of β-catenin in the nucleus, where it acts to inhibit a pluripotency repressor TCF3, and in the cytoplasm. Cytoplasmic β-catenin is free to take part in the formation and stabilization of cell-cell adherens junctions (Tian et al., 2011). Upon withdrawal of chiron, the re-establishment of β-catenin destruction complex lead to destabilization of these adherers junctions which result

in cells that are unable to form cell-cell contact (Andersson-Rolf et al., 2017; Wray et al., 2010). Furthermore, there are changes in RNA expression of cell-cell adhesion molecules, including the down-regulation of Pecam1 and Icam1 and the up-regulation of Cldn7 (Kalkan et al., 2017). It follows that ES cells should form compact colonies of tightly joined cells, whereas colonies formed of exiting cells should be less associated. This has implications in cell motility and cell size. If cells are are forming more cell-cell adhesion, then the cells will be more constricted and less motile than cells with fewer cell-cell adhesion. The number of cell-cell adhesion should also effect the the size of the cells. Thus I examine how cell size, motility, and cell-cycle vary over time during the exit process at the single cell resolution (Fig. 1.15).

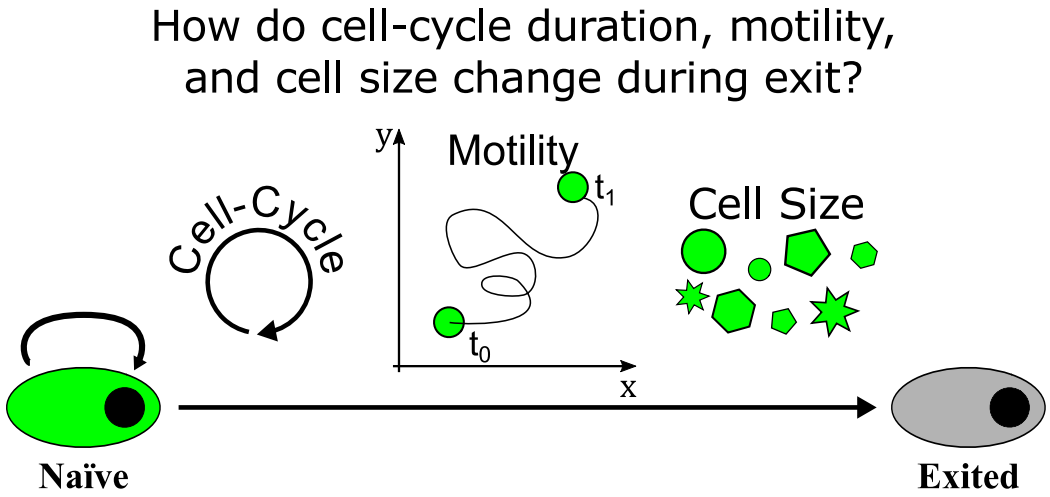


Fig. 1.15 ES cell colonies have compact morphologies while colonies formed during the exit process exhibit a sparse morphology. The withdrawal of the 2i results in the reactivation of Erk signaling and of the β-catenin destruction complex. These signaling pathways play rolls the cell-cycle, motility, and cell-cell adhesion. So how are these processes affected during the exit process?

How do single cells down-regulate Rex1-GFPd2?

The single-cell kinetics of Rex1-GFPd2 down-regulation upon exit from naïve pluripotency are not known (Martello and Smith, 2014). In 2i, bulk-culture assays, like flow cytometry and clonogenecity assays, it is seen that Rex1-GFPd2 is down-regulated over a period of days (Kalkan et al., 2017; Mulas et al., 2017). However, Fig 1.11 and Fig 1.12 shows how these data are unable to discern how an ensemble of individual cells generate such behavior. In fact, there are many trajectories which single-cells can travel along during the down-regulation of Rex1-GFPd2 which could result in the same population kinetics observed.

Yet it is certain from these bulk-culture assays that individual cells, within the population in the dish, are exiting naïve pluripotency at different times (Kalkan et al., 2017; Mulas et al., 2017).

In the serum/Lif culture system, the expression of key pluripotency factors, such as Nanog and Rex1, exist in a metastable state (Chambers et al., 2007; Chickarmane et al., 2006; Glauche et al., 2010b; Kalmar et al., 2009; Toyooka et al., 2008). The steady-state distribution of Nanog is bimodal, with a dominant Nanog high peak and a smaller Nanog low peak. Long-term single-cell imaging and tracking show that this bimodality is a result of stochastic switching between a Nanog high and a Nanog low state (Herberg et al., 2016). This is thought to result from a negative feed back loop in the core circuitry induced by the inductive signals in serum/Lif (Herberg et al., 2016).

Although the 2i and serum/Lif system generate different expression patterns in the naïve circuitry, both systems stimulate the same network. In that vein, universal lessons can be learned from each. First, the serum/Lif system can give insight into how Rex1 expression is regulated. Herberg et al., 2016, have shown that the transition between the high and low states of Nanog and Rex1 occurs by stochastic switching, indicating that a sudden switch is likely to be expected upon exit. Second, the 2i system recapitulates the expression of naïve factors in the epiblast. Smith, 2017, has proposed the phased progression model of pluripotency, where cells incrementally progress from naïve, to formative (the executive phase of pluripotency), to primed pluripotency. For ES cell behavior to be consistent with these data and this hypothesis, Rex1-GFPd2 down-regulation should proceed by a single-step switch, with individual cells shutting-off Rex1-GFPd2 at different times over a period of days.

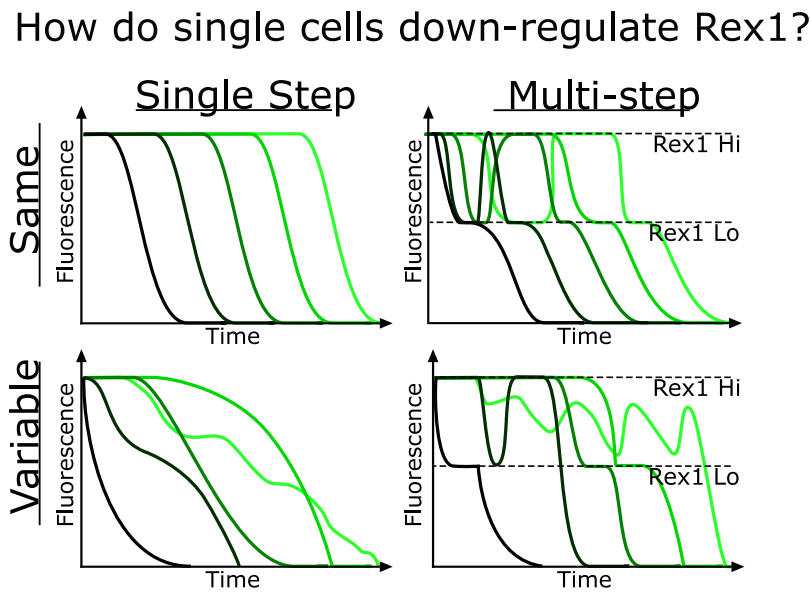


Fig. 1.16 The single-cell kinetics of Rex1-GFPd2 down-regulation upon exit from naïve pluripotency are not currently known. However, it is know that individual ES cells exit at different time over a period of days. So, how does the ensemble of cells behave during exit: Do cell proceed to exit in a single step or multi-step process? When they ultimately exit, do all cells proceed by the same kinetics or are their kinetics varied?

However, during the transition to exit from naïve pluripotency from 2i, Rex1-GFPd2 down-regulation kinetics can be parsed into four potential classes (Fig. 1.16). The first question for Rex1-GFPd2 down-regulation asks if the process occurs in a single step, consistent with the phased progression model, or by reversibly transitioning though multiple states. The second question asks if all of the cells are behaving in the same manner. Individual cells have the ability to down-regulate the reporter with the same kinetics or with different kinetics. Finally, embedded within these classes of ensemble behavior are the single-cell kinetics. When cells ultimately down-regulate Rex1-GFPd2, is it a gradual silencing of the gene, a sudden switch, or is it perhaps something more exotic (Fig. 1.16)?

How do ES cells divide during the exit from naïve pluripotency: symmetric, asymmetric, or mixed?

In the adult body, tissues and organs must be sustained over the lifespan of the organism. This requires the generation of new, tissue specific cells to replace cells lost during normal turn over (i.e. homeostasis) and during regeneration, when a tissue must replace cells lost during some form of damage (e.g. a cut on the skin). These processes are managed by tissue specific populations of adult stem cells (Simons and Clevers, 2011). A micro-

environment, called the niche, is formed by cells in the tissue which provide critical signals for adult stem cell self-renewal. Usually, cells that do not receive signals from the niche begin to differentiate.

In homeostasis and regeneration, it is imperative to maintain these population of stem cells. A disproportionate level of differentiation results in the loss of the stem cell population and the inability for long-term tissue maintenance. If there is not enough differentiation, then the tissue will not be sufficiently populated with cells to carry out the normal functions of the tissue. Adult stem cell populations are maintained by using several methods of divisions. This includes the use of asynchronous symmetric divisions, and asymmetric divisions (Fig. 1.17) (Simons and Clevers, 2011). The typical source of asymmetric division is from when one daughter detaches from the niche and thus stops receiving self-renewal signals, so it subsequently differentiates.

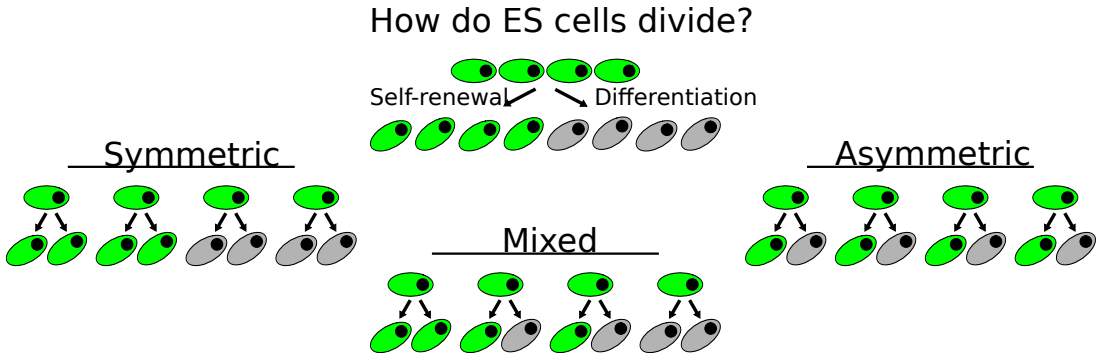


Fig. 1.17 The symmetry of a cell divisions is determined by the fate of the daughter cells. In differentiation, cells can either divide symmetrically or asymmetrically. A symmetric divisions results in two daughters of the same identity. Asymmetric divisions result in daughter cells with different identities. Asymmetric divisions are either extrinsically or intrinsically regulated. Extrinsic regulation requires the two daughter cells to encounter different environments. Intrinsic regulation requires a polarized cell to divide along the axis of polarization such that cell fate determinants are differentially allocated into the daughter cells. What type of divisions do ES cells, are non-polarized cell lacking an *in vitro* niche, perform: symmetric, asymmetric, or a combination of the two? (Adapted from Morrison and Kimble, 2006)

But what happens in the case of ES cells differentiating *in vitro*? ES cell recapitulate a very early stage of development. Morrison and Kimble, 2006, note that during neural and epidermal development, stem cells primarily divide symmetrically to expand the stem-cell pool while later, during late gestation, asymmetric divisions are used to expand differentiated cell population. ES cells, and the pre-implantation Epiblast for that matter, are non-homeostatic tissues. Once self-renewal signals are withdrawn from ES cells, they are all destined to

differentiate or die. They have no need to maintain the original ES cell population like that of the adult tissue. So there is no seeming necessity for asymmetric to occur.

The requirement for extrinsic regulation of asymmetric divisions are a localized source of self-renewal signals. In the culture dish, there are no distinct niche structures creating microenvironments conducive to the symmetry breaking found *in vivo*. So this makes extrinsically regulated differentiation seem unlikely.

The Case for intrinsically regulated asymmetric divisions seems as unlikely. ES cells on gelatin do not exhibit polarization (Shahbazi et al., 2017). However, on matrigel, ES cells are able to initiate polarization, but fail to undergo lumenogenesis, and become disordered. Furthermore, ES cells on matrigel and gelatin exhibit the same expression levels of naïve factors. In fact, the naive gene expression network is dismantled at lumenogenesis on matrigel, thus uncoupling any functional form of ES cell polarization from the naïve circuitry.

Does the asynchronicity of exit arise within individual colonies or only within the population?

Upon withdrawal of self-renewal signals, ES cells dwell in a recoverable transition state until exit occurs (Betschinger et al., 2013; Kalkan et al., 2017; Leeb et al., 2014). This decision to exit is made asynchronously, where cells exit at different times over a period of days. The asynchronicity of differentiation and lineage specification has been documented both *in vivo* (Acampora et al., 2016; Saiz et al., 2016) and *in vitro* (Betschinger et al., 2013; Kalkan et al., 2017; Leeb et al., 2014; Lowell et al., 2006; Zhou et al., 2013). However, the population strata at which asynchronicity arises is unknown.

Snap-shot data does not preserve the relationship cells have to one another. So, while it is possible to determine that single cells are exiting asynchronously within a culture dish, it is not possible to indefinitely determine if asynchronous exit scales down to a colony generated by a single cell. That is, does a colony generated by a single cell exhibit synchronous or asynchronous exit?

This question can be answered by determining how individual colonies generate asynchronous populations of single cells. Individual colonies can generate a population of asynchronously exiting single cells by using different combinations of cell divisions and intraclonal synchronicity. First consider the case where individual colonies exhibit synchronous exit (Fig. 1.18). This is where, like clockwork, all the cells initiate differentiation. If cells only diving symmetrically during exit, then all of the cells withing a colony will exit simultaneously. Individual colonies would need to exit at different time to generate a

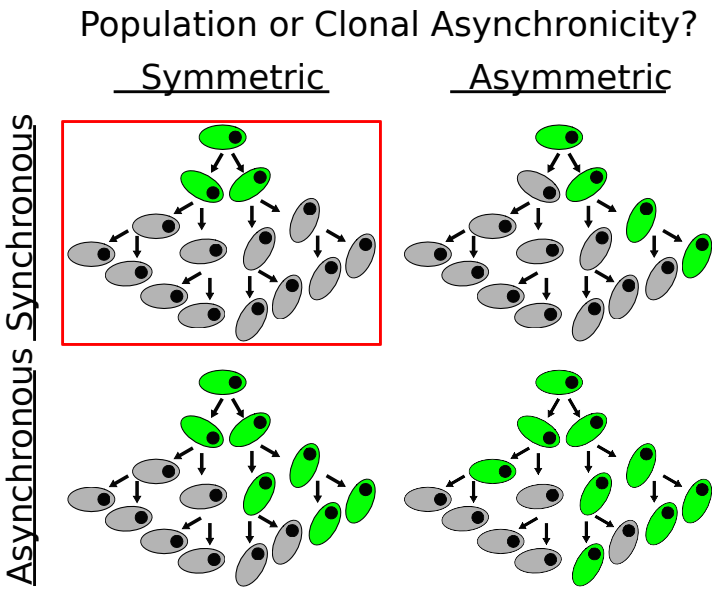


Fig. 1.18 Individual colonies can give rise to a population of asynchronously exiting single cells by using different cell division types and different intraclonal synchronicity. ES cells that employ symmetric divisions with synchronous intraclonal dynamics form colonies that exit simultaneously. This would require many colonies to exit at different to generate population asynchronicity. In the case where cells use asymmetric divisions and synchronous intraclonal dynamics, population asynchronicity is rapidly diluted upon successive cell divisions. Population level asynchronicity will arise from cells that use asynchronous intraclonal dynamics, regardless of division type.

population of individual cells exiting asynchronously. Asymmetric divisions will give rise to both clonal and population asynchronicity. However, the clonal asynchronicity will appear to vanish as the population of recoverable cells becomes negligible relative to the number of exited cells. If cells use asynchronous intraclonal dynamics, then the population of individual cell will be inherently asynchronous, regardless of division type.

Determining the origin of the population level asynchronicity is important in understanding the programming, or lack there of, in the dissolution of the naïve pluripotency network. If the intraclonal dynamics are asynchronous, then this would imply that the decision to exit is made within the lifetime of an individual cell. Meaning that exit occurs in a probabilistic fashion dictated simply by the time spent under factor withdrawal conditions. However, synchronous exit, where cells several generations apart down-regulate Rex1-GFPd2 simultaneously, would imply that an ancestral cell has made an instructive decision that is inherited. The implication being a deterministic program induces the dissolution of the naïve circuitry.

Chapter 2

Materials and Methods

2.1 Tissue Culture and Materials

2.1.1 Embryonic Stem Cell Lines

E14-IVc. WT mouse embryonic stem (ES) cell line was derived in-house from embryos using previously described protocols (Nichols and Smith, 2009).

Rex1-GFPd2. The Rex1-GFPd2 ES cell line was derived from E14-IVc cells. This cell line was generated by replacing one allele of the Rex1 coding region with a destabilized green fluorescent protein (GFPd2) linked to a blasticidin-resistance cassette by homologous recombination (Kalkan et al., 2017).

Gap43-mCherry. The Gap43-mCherry reporter ES cell line was derived from E14-IVc ES cell line. To generate the reporter, the palmitoylation sequence of Gap43 was introduced to the N-terminus of the mCherry sequence via PCR amplification. The chimeric Gap43-mCherry PCR product was digested with Sal1 and XhoI and was subsequently ligated to Sal1/XhoI digested pEntr2B vector. The ligation product was transformed into competent *E. coli* strain DH5α and selected for by kanamycin resistance. Individual clones were picked and cultured. Plasmid DNA was extracted from the culture and insertion of the PCR product was verified by Sal1/XhoI restriction digest and gel electrophoresis. The resulting plasmid, pEntr2B-Gap43-mCherry, was used as the intermediate vector for subsequent Gateway cloning. 100ng pEntr2B-Gap43-mCherry plasmid was mixed with 100 ng destination vector, pPB-CAG-pA-pgk-hph, together with LR Clonase reaction mix following the manufacturer's protocols (ThermoFisher 11791). The reaction was incubated at room temperature for 1

hour at which point Proteinase K was added to stop the reaction prior to transformation with DH5α *E. coli*. Clones were selected for by plating on ampicillin-containing agar plates. Ampicillin resistant clones were picked and expanded for DNA purification. Insertion of the chimeric protein DNA sequence into the Gateway destination vector was verified by restriction digestion with HindIII prior to sequencing. The resulting plasmid was named pPB-CAG-Gap43-mCherry-pgk-hph.

GAP43-mCherry was stably expressed in ES cells by co-transfection of pPB-CAG-Gap43-mCherry-pgk-hph with PB transposase (4:1 ratio) using Mirus TransIT LT1 lipofection reagent following the manufacturer's protocols (MIR 2300) . Transfected cells were selected using 200 μg/ml hygromycin in 2i + 10% FBS for a minimum of 7 days for stable PB integration. Individual clones were picked for brightness and expanded.

Table 2.1 Table of primers for Gap43-mCherry cloning

Primer	Sequence
SalI-Kozak-Gap43-Palm Forward	AAAAGTCGACTGCCACCACCATGCTGTGCTGTATG AGAAGAACCAAACAGGTTGAAAAGAATGATGAGGA CCAAAAGATCATGGTGAGCAAGGGCGAGGAGGATA
XhoI-mCherry Reverse	TTTTCTCGAGTTACTTGTACAGCTCGTCCATGCCG
iGap43 Forward	CGGTATCGATAAGCTTGCCACCACCATGCTGTGC
imCherry Reverse	TGATTACGCCAAGCTTTTACTTGTACAGCTCGTC
pEntinSeq Forward	TTTCTACAAACTCTTCCTGTTAGTT
pEntinSeq Reverse	GTAACATCAGAGATTTTGAGACAC
pCAG seq Forward	CTCCTGGGCAACGTGCTGGT
pCAG seq Reverse	TGTTGTGTGGAATTGTGAGC

Rex1-GFPd2/Gap43-mCherry.The Rex1-GFPd2/Gap43-mCherry ES cell line is a sub-clone of the Rex1-GFPd2 ES cell line. The Gap43-mCherry reporter was transfected in the same fashion as the Gap43-mCherry ES cell line. The kinetics of Rex1-GFPd2 down-regulation was validated against the parental line (Shown in Fig. 4.3).

2.1.2 Media and Reagents

2.1.2.1 Cytokines and Inhibitors

Lif, used to supplement medium at 10 ng/mL (Department of Biochemistry, University of Cambridge). GSK3 inhibitor Chiron, used to supplement medium at 3μM (acquired from an unnamed collaborator). MEK inhibitor PD03, used to supplement medium at 1μM (acquired from an unnamed collaborator).

2.1.2.2 Culture Media

Serum Basal Medium. GMEM (Sigma-Aldrich, G5154) supplemented with 10% FCS (v/v) (Sigma-Aldrich), 1x NEAA (Life Technologies, 11140-050), 1mM sodium pyruvate (Life Technologies, 11360-070), 1mM L-Glutamine (Life Technologies, 25030-081).

Lif & PD03 in Serum (LPS). Serum basal medium supplemented with Lif and PD03.

Serum-based Freezing Medium. LPS with 20% DMSO (v/v) (Sigma-Aldrich, 472301).

Serum-based Wash Buffer. Serum basal medium.

N2B27 Basal Medium. Equal volumes of DMEM/F12 (Life Technologies, 21331-020) and Neurobasal medium (Gibco, 21103-049) supplemented with 1% (v/v) B27 with RA (Gibco, 17504-044), 0.5% (v/v) N2 (made in-house), 0.1 mM b-mercaptoethanol (Life Technologies, 31350-010), and 2mM L-Glutamine (Life Technologies, 25030-081).

2i. N2B27 supplemented with Chiron and PD03.

N2B27-based Freezing Medium. 2i with 20% DMSO (v/v) (Sigma-Aldrich, 472301).

N2B27-based Wash Buffer. DMEM/F12 (Life Technologies, 21331-020) with 0.03% (v/v) BSA (Life Technologies, 15260-037).

2.1.2.3 Miscellaneous

PBS, Dulbecco's Phosphate Buffered Saline (Sigma-Aldrich, D8537). Accutase, ESGRO Complete Accutase (Millipore, SF006). Trypsin in chick serum, 0.025% Trypsin (Invitrogen, R001100), 1% chicken serum (Sigma-Aldrich) in PBS. Gelatin, 0.1% (w/v) Gelatin (Sigma-Aldrich, G1890) in PBS. Laminin, 5 ng/mL Laminin (Sigma-Aldrich, L2020) in PBS. Hoechst33342, 10 mg/mL stock solution, 1 mg/mL working solution (ThermoFisher, H3570).

2.1.3 Tissue Culture

All cells were manipulated under laminar flow hoods and grown in humidified incubators kept at 37°C in 7.5% CO_2. All plates were gelatinized for at least 20 minutes with 3 mL of gelatin in PBS and allowed to dry for an additional 5 minutes post aspiration.

Bulk Culture Maintenance. Cells were grown in 3 mL of 2i or LPS on flat bottom 6-well tissue culture plates (Falcon, 353046) on gelatin. Cell were fed every other day and passaged every 2-3 days.

Thawing Cell Lines. Cell were thawed in a 37°C water bath and diluted with 10 mL of wash buffer. The cell suspension was centrifuged for 3 minutes at 1.3×10^3 rpm. The pellet was resuspended in 2i or LPS and cells were plated on the order of 1×10^4cell/cm^2 (o(10^5) cells per well).

Passaging Cell Lines.

- **N2B27-based medium**. Medium was aspirated and 500 μL of Accutase was added to the well. After the cells had detached from the well, no greater than 5 minutes, 2 mL of N2B27-based wash buffer was added to the well. Cells were then mechanically separated via pipetting. The cell suspension was centrifuged for 3 minutes at 1.3×10^3 rpm. The pellet was resuspended in 2i and cells were plated on the order of 1×10^4cell/cm^2 (o(10^5) cells per well).
- **Serum-based medium**. Medium was aspirated, cells were washed with 3 mL of PBS, and 500 μL of trypsin was applied to the well. After the cells had detached from the well, no greater than 3 minutes, 2 mL of Serum-based wash buffer was added to the well. Cells were then mechanically separated via pipetting. The cell suspension was centrifuged for 3 minutes at 1.3×10^3 rpm. The pellet was resuspended in LPS and cells were plated on the order of 1×10^4cell/cm^2 (o(10^5) cells per well).

Freezing Cell Lines. Once cells had been resuspended after passaging, they could be frozen. This was accomplished by mixing equal parts of cell suspension with Freezing Medium. This resulted in a final DMSO concentration of 10% (v/v).

2.2 Kinetics Assays

2i. Cells for all assays were maintained in humidified incubators kept at 37°C in 7.5% CO_2 and 5% O_2. Low O_2 was used for improved cell viability during differentiation and recovery (Mazumdar et al., 2010). Assays were conducted in Dubelco 6-well plates coated with laminin. Laminin coatings were allowed to set for at least 12 hours prior to plating cells.

LPS. Cells for all assays were maintained in humidified incubators kept at 37°C in 7.5% CO_2 and 20% O_2. Assays were conducted in Dubelco 6-well plates coated with gelatin.

2.2.1 Population-Level Recovery Kinetics

2i. Rex1-GFPd2 cells were plated on laminin at a density of 5000 cells/10 cm^2 in 2i. After 12 hours, the medium was aspirated and replaced with N2B27. 2i was serially reapplied for the different time-points: 16, 24, 32, 40, and 48 hours post withdrawal. A positive control was maintained in 2i throughout the withdrawal time-course. Cells were allowed to recover for at least 24 hours in 2i, at which point flow cytometry was performed on all samples (performed in technical duplicate).

LPS. Rex1-GFPd2 cells were plated on gelatin at a density of 5000 cells/10 cm^2 in LPS. After 12 hours, the medium was aspirated and replaced with Serum basal medium. LPS was serially reapplied for the different time-points: 24, 48, 72, and 96 hours post withdrawal. A positive control was maintained in LPS throughout the withdrawal time-course. Cells were allowed to recover for at least 24 hours in LPS, at which point flow cytometry was performed on all samples (performed in technical duplicate).

All samples were analyzed with the CyAn ADP Analyser (Beckman Coulter), with 488nm laser and fluorescein isothiocyanate (FITC) 530/40nm and phycoerythrin (PE) 575/25nm band pass filters. Flow profiles were analyzed and quantified with Summit V4.3 (Dako Colorado, Inc.).

2.2.2 Clonal Recovery Kinetics

2i. Rex1-GFPd2 cells were plated on laminin at single-cell density (500 cells/ 10 cm^2) in 2i. After 12 hours, the medium was aspirated and replaced with N2B27. 2i was serially reapplied for the different time-points: 16, 24, 32, 40, and 48 hours post withdrawal. A positive control was maintained in 2i throughout the withdrawal time-course. Cells were then allowed to recover for at least 24 hours in 2i. Each sample was stained for nuclei (Hoechst), alkaline phosphatase (AP), and Rex1-GFPd2 (GFP) and imaged (performed in technical duplicate).

LPS. Rex1-GFPd2 cells were plated on gelatin at single-cell density (500 cells/ 10 cm^2) in LPS. After 12 hours, the medium was aspirated and replaced with Serum basal medium. LPS was serially reapplied for the different time-points: 24, 48, 72, and 96 hours post withdrawal.

A positive control was maintained in LPS throughout the withdrawal time-course. Cells were then allowed to recover for at least 24 hours in LPS. Each sample was stained for nuclei (Hoechst), alkaline phosphatase (AP), and Rex1-GFPd2 (GFP) and imaged (performed in technical duplicate).

All plates were scanned using an Olympus IX51, Prior stage and Lumen light source, with Cell Sens software, using a 10x objective. Images were analyzed using ImageJ (Rasband et al., 2014).

2.2.3 Proliferation Kinetics

2i. Rex1-GFPd2 cells were plated on laminin at single-cell density (500 cells/ 10 cm^2) in 2i. After 12 hours, the medium was aspirated and replaced with N2B27. At 16, 24, 32, 40, and 48 hours post withdrawal, colonies were stained for nuclei with Hoechst and imaged at high-resolution. Positive control cells, maintained in 2i, were stained and imaged at each time-point in parallel (performed in technical triplicate).

N2B27. Rex1-GFPd2 cells were plated on gelatin at single-cell density (500 cells/ 10 cm^2) in LPS. After 12 hours, the medium was aspirated and replaced with serum basal medium. At 24, 48, 72, and 96 hours post withdrawal, colonies were stained for nuclei with Hoechst and imaged at high-resolution. Positive control cells, maintained in LPS, were stained and imaged at each time-point in parallel (performed in technical triplicate).

Plates were scanned using an Olympus IX51, Prior stage and Lumen light source, with Cell Sens software, using a 10x objective.

Quantification of image for doubling times are explained in detail in Sec. 3.3.1. All images were analyzed using ImageJ Rasband et al., 2014. Regression analysis and quantification of number of cells per colony were performed with MATLAB 2016a (MathWorks, Inc.) and doubling times were calculated with Prism 7.00 (GraphPad Software, Inc.).

2.3 Long-Term Single-Cell Imaging, Tracking, and Analysis

The transition to exit from naïve pluripotency was monitored by the Rex1-GFPd2 reporter system, which is inactivated upon exit (Kalkan et al., 2017; Mulas et al., 2017). Culture conditions that provided homogeneous populations were used in all assays to mitigate heterogeneity from confounding synchronicity information. Bulk-culture assays were performed to quantify the population behavior of exit from naïve pluripotency and proliferation. The time scale of exit from naïve pluripotency, as determined by these assays, was used to set the duration of long-term single-cell imaging.

Quantitative long-term single-cell imaging over many generations was used to gather kinetic data for cell position, cell size, and Rex1-GFPd2 fluorescence. Confocal imaging was conducted using the Rex1-GFPd2 ES cell line with a fluorescent-protein tracking reporter. The plasma membranes of the ES cells were labeled with an mCherry N-terminally tagged with the palmitoylation sequence from the membrane associated protein Gap43. Semi-automated tracking was conducted using bespoke tracking and analysis software (Blanchard et al., 2009). Images of cells were manually identified and segmented, while collation of raw kinetics and genealogical data was automated. The raw data were processed and analyzed using a suite of custom MATLAB functions. Clonal lineage trees were constructed and used to conduct genealogical analyzes for different cellular attributes: size, motility, cell-cycle duration, and Rex1-GFPd2 expression level. The Rex1-GFPd2 signal in differentiating cells was controlled for by Rex1-GFPd2 ES cells in 2i as the positive control and WT ES cells in 2i as the negative control. 2i ES cells were also used as the control for cell size, motility, and cell-cycle duration. These data were validated against the bulk-culture data for exit kinetics and doubling time for both 2i ES cells and N2B27 cells.

2.3.1 Confocal Microscopy

Cells were plated on laminin at single-cell density (500 cells/ 10 cm^2) in 2i and placed in humidified incubators maintained at 37°C in 7.5% CO_2 and 5% O_2. After 12 hours, the medium was aspirated and replaced with N2B27 for differentiation conditions or 2i for control conditions. The culture dish was returned to the 37°C humidified 7% CO_2 and 5% O_2 incubator for one hour to allow for medium to equilibrate prior to imaging.

Time-lapse imaging was performed using a Leica SP5 confocal microscope with the Leica 20x 0.7NA dry lens (Leica, 506513). Seven adjacent 3.7 nm z-stacks were acquired at multiple positions for GFP (488nm excitation with Ar laser, 500-550nm emission detected by HyD2 at 150% gain and AOTF 15), mCherry (594nm excitation with HeNe laser, 600-630nm emission detected by HyD4 at 100% gain and AOTF 12), and bright field at 30 minute intervals for 72 hours. Laser power was set at 30% and pinhole set at 2 Airy units. The microscope was enclosed in a temperature controlled chamber, 'The Box' heat by 'The Cube' (Life Imaging Sciences), and maintained at 37°C in humidified 7% CO_2 and 5% O_2.

2.3.2 Tracking Software

Bespoke semi-automated tracking (Tracer) and analysis (Otracks) software was written and modified by Blanchard et al., 2009, in IDL (Harris Geospatial Solutions). The Tracking and analysis software is available upon request from Blanchard et al., 2009. The software was run on OS X Mavericks (Apple Inc.), using the VMware Workstation 10.0.1 (VMware, Inc.), hosted on Windows 8.1 Pro (Microsoft).

2.3.3 Computing and Statistics

Prism 7.00 (GraphPad Software, Inc.) was used to perform two-way ANOVA using Holm-Sidak post-test to correct for multi-way comparisons. Normality tests were conducted on data sets prior to AOVA (see App A). Unless otherwise stated, MATLAB 2016a (MathWorks, Inc.) was used to conduct Student's t-Tests and to calculate Pearson's correlation coefficients.

Novel software was constructed and implement in MATLAB 2016a to handle and analyze the raw data produced by Tracer and Otracks (Blanchard et al., 2009). This consists of a library of functions that constructs and manipulates data-embedded lineage-trees (see Sec. 4.4 for details). The folder structure for this library is detailed here and in Fig. 2.1.

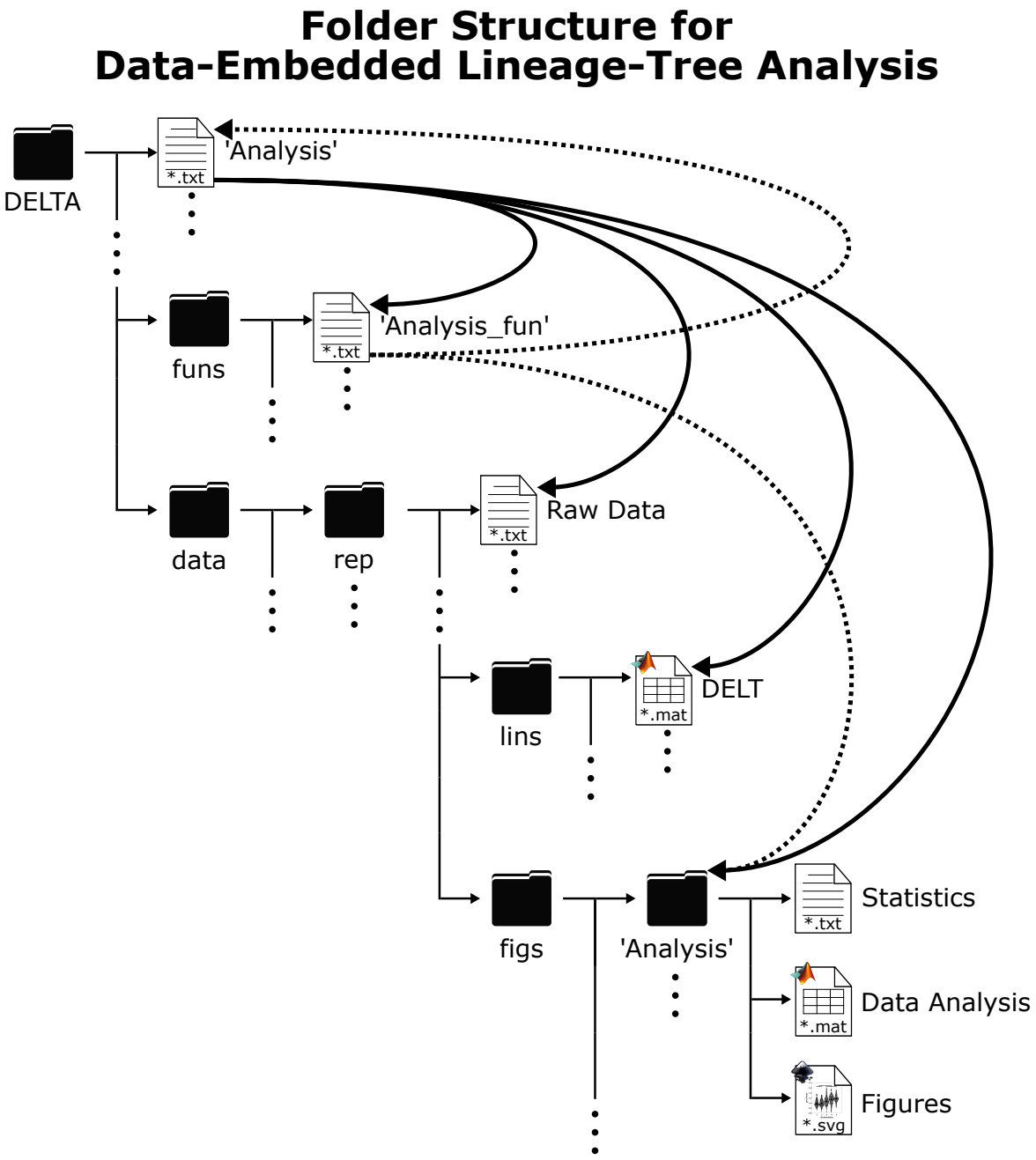


Fig. 2.1 The folder hierarchy for Data-Embedded Lineage-Tree Analysis (DELTA) provides the foundation for systemic and reproducible analysis of the raw lineage data output from Otracks. DELTA is performed on this set of raw '.txt' files generated by Otracks using a suite of custom 'Analysis' scripts and 'Analysis_funs' functions. The 'Analysis' scripts work by calling the 'Analysis_funs' which in turn pass information from the Data-Embedded Lineage-Trees (DELT) through a series of data and statistical analysis protocols. These protocols perform lineage reconstruction, data analysis, statistical analysis, and figure generation. The output of these protocols are summarized in the 'Analysis' folders which consist of the aforementioned Data Analysis, Statistics, and the resulting Figures.

The data-embedded lineage-tree analysis (DELTA) folder houses all of the analytical methods in the forms of 'Analysis.txt' scripts. In total, 'Analysis.txt' scripts have been written for the following analyses:

- Data-Embedded Lineage-Tree reconstruction
- Non-arbitrary signal thresholding
- Single-cell attribute visualization: Rex1-GFPd2 levels, cell-cycle, motility, and cell size.
- Generational Kinetics: cell-cycle, motility, and cell size.
- Genealogical relationships: Rex1-GFPd2 levels and cell-cycle

Details for the individual analyzes can be found in the remainder of this chapter and in Chapter 4 and the results of these analyzes can be found in Chapter 5. The 'Analysis' scripts have access to both 'funs' and 'data' and all subfolders.

'DELTA\funs' houses subroutine scripts 'Analysis_funs.txt' called by the 'Analysis.txt' script in 'DELTA'. These scripts consist of functions used to do batch processing, plotting, data manipulation, and statistical analysis. The 'Analysis_funs.txt' scripts have access to call other scripts in 'DELTA\funs', feed information back to the 'Analysis.txt' scripts, and have access to save data to 'Delta\data\rep\figs\Analysis'.

The 'DELTA\data' folder house subfolders for individual biological replicates 'DELTA\data\rep'. In this study two biological replicates were performed, thus, the 'DELTA\data' folder houses two subfolders. The 'DELTA\data\rep' folders store the 'Raw Data.txt' files from Otracks, the 'DELTA\data\rep\lins' folder, and the 'DELTA\data\rep\figs' folder. The first 'Analysis.txt' run is 'EXTRACT.txt'. This reconstructs lineages using the 'Raw Data.txt' files, and stores them as data-embedded lineage-trees ('DELT.mat') in the form of MATLAB structures. This structure is detailed in Sec. 4.4. The 'DELT.mat' files are data used for all analysis.

The 'DELTA\data\rep\figs' folder, a slight misnomer, houses subfolders for each 'Analysis.txt' performed. The 'DELTA\data\rep\figs\Analysis' folders are the ones that store the 'Statistics.txt' and 'Data Analysis.mat' generated and called by the 'Analysis.txt' and 'Analysis_fun.txt' scripts, and, indeed, 'Figures.svg'.

2.3.4 Nonarbitrary Signal Thresholding

Rex1-GFPd2 signal thresholding was performed using the positive and negative controls from the long-term single-cell imaging and tracking. For each replicate, all timepoints from

all cells (see Table 4.1) were individually projected onto the fluorescence axis to generate the distribution of expression levels for the control (Fig. 2.2). For the positive threshold, 6,141 data points from 405 cells from 17 clones were used for replicate 1 and 2,438 data points from 154 cells from 18 clones were used for replicate 2. For the negative threshold, 664 data points from 56 cells from 6 clones were used for replicate 1 and 1,258 data points from 52 cells from 6 clones were used for replicate 2.

These data were transformed by $\log_{10}$ and a normal distribution was fit to the resulting histogram. The lower bound for the positive control was determined by generating a one-tailed lower 95% confidence interval for the positive control, where anything below the critical value was considered Rex1-GFPd2 negative. Similarly, a one-tailed upper 95% confidence interval was calculated for the negative control, where anything above the critical value was considered Rex1-GFPd2 positive.

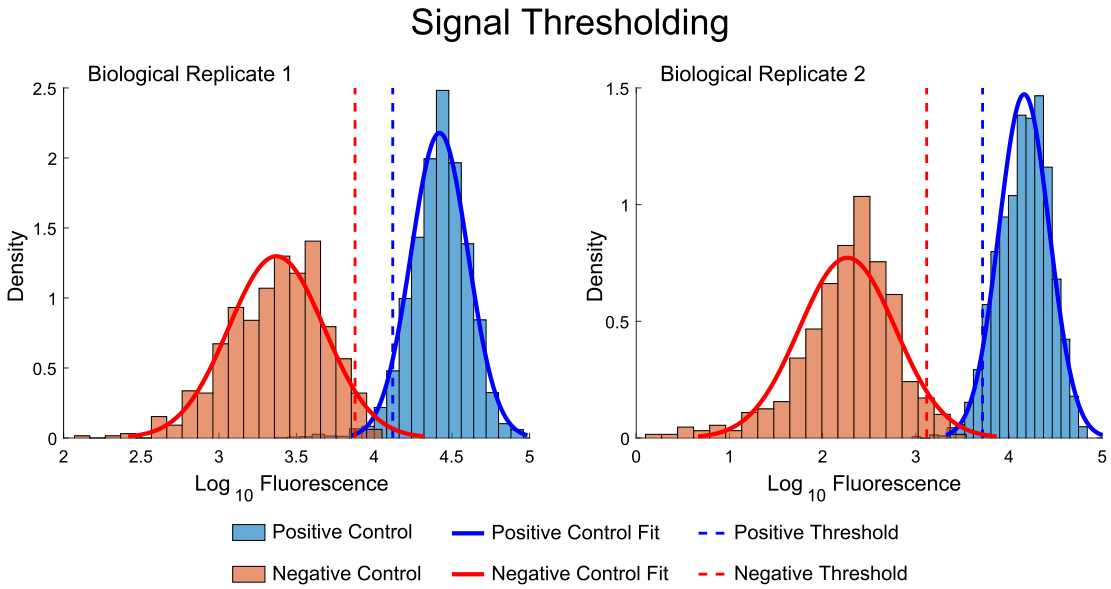


Fig. 2.2 Signal thresholding for Rex1-GFPd2. The on/off threshold for Rex1-GFPd2 was determined using the positive control (Rex1-GFPd2 ES cells in 2i) and negative control (E14-IVc ES cells in 2i). Histograms for each conditions were generated by projecting all data-points from all cells within an experiment onto the fluorescence axis. A normal distributions were fit to the resulting histograms and the resulting one-tailed 95% confidence interval from the controls were used to set limits for the GFPd2 positive and negative status of a cell. If using the positive threshold (blue dashed line), anything below the critical value was considered negative. Likewise, if using the negative threshold (red dashed line), anything above the critical value was considered positive.

2.3.5 Time-Shifting and Ordering

Time-Shifting of genetic trajectories were conducted to examine common behaviors that are disparate in time. Lineage alignment was performed by determining the time the first cell in a lineage crossed the calculated threshold (Fig. 2.3 A). The new origin for that lineage was set at the point the trajectory crossed the threshold and all lineages were superimposed (Fig. 2.3 B). Individual origins were then shifted to a global origin (Fig. 2.3 C).

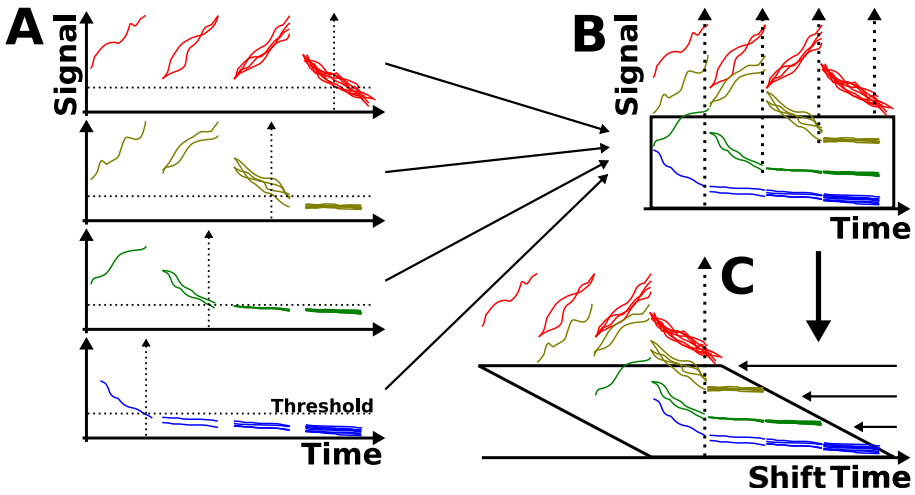


Fig. 2.3 Time-Shifting of Genetic Trajectories. Genetic trajectories for individual colonies were shifted relative to one another to elucidate common patterns. (A) Time-shifting was performed by first establishing an on/off threshold (horizontal dotted line) for Rex1-GFPd2 (detailed in Sec. 2.3.4). A new origin was set at the time which the first cell of a lineage crossed the calculated threshold (Vertical dotted line). (B) All lineages from the experiment were superimposed and then (C) shifted to a common origin.

As part of the cell-cycle analysis, branch-traces of genetic trajectories were progressively ordered by the time of division (Fig. 2.4). Branch-traces of lineages were constructed by root-to-tip traces for each terminal branch. These were then progressively ordered according to the times at which divisions occurred.

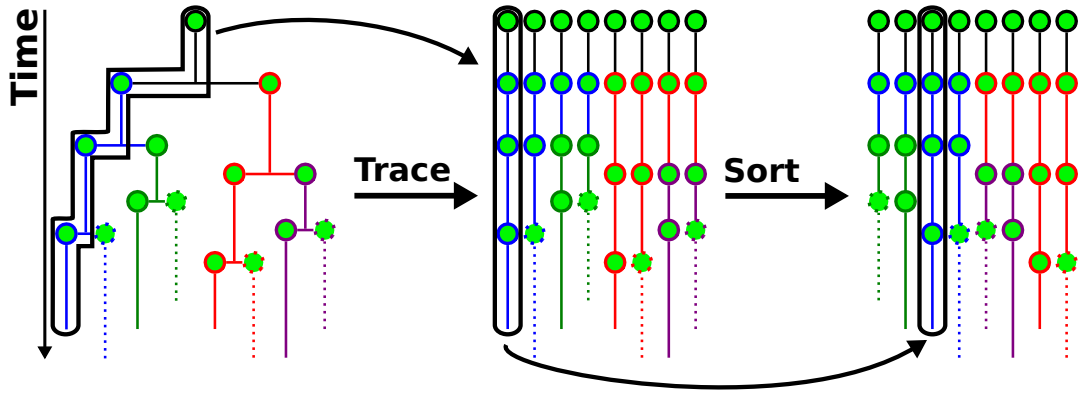


Fig. 2.4 Branch-Trace Ordering Time of Divisions. Patterns of Rex1-GFPd2 expression relative to times of divisions were examined by creating and progressively ordering branch-traces. Each branch of a lineage-tree was traced from root-to-tip to generate branch-traces for all lineages. These were then progressively ordered by the times at which divisions occurred along the branch-trace. Individual trajectories were plotted in heat-maps according to this order (see Sec. 5.13).

2.3.6 Approximation of Growth Curves and Exit Kinetics

One recurring issue was the loss of traceable cells in large colonies. This has implications in the accuracy of calculating the growth rate and the exit kinetics. This was mitigated by using empirical data to approximate their values.

Growth curves were generated by plotting the time-course of the total number of cells present (Fig 2.5 A). However this number becomes inaccurate if cells become untraceable. The clonal growth curves were approximated by creating individual growth curves for the branches where cells became untraceable. Generational distributions of cell-cycle durations were used to randomly generate the time a cell would next divide (Fig 2.5 B,C). This process was repeated until all time-points were completed. These were then added to the clonal growth curve.

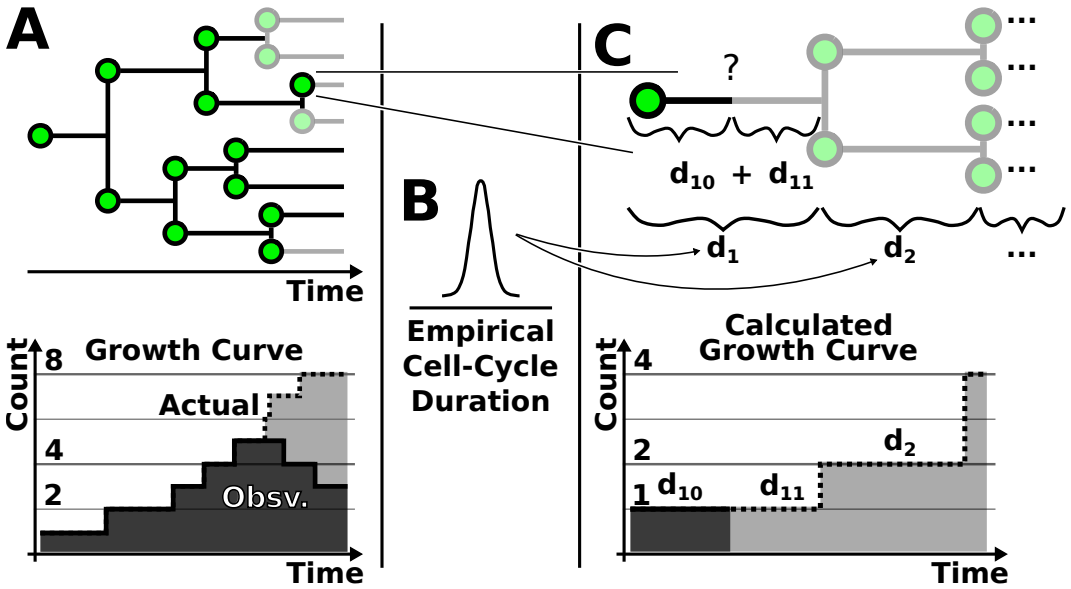


Fig. 2.5 Approximation of Single-Cell Growth Curves. (A) At certain times, cells of a lineage tree become untraceable (top, grayed branches), which results in inaccurate growth curves (bottom). (B) Thus, growth curves were estimated using the empirical distribution of cell-cycle duration. (C) A cell which is lost will be track for a time d_{10}. However, this is only part of the life of the cell d_1. The remainder of the cells life-span,d_{11} is determined from an empirical distribution ($\mathrm{d}_1 = N(\mu, \sigma^2)$). At this new time, a cell division has occurred, so an additional cell was added to the total cell count. All subsequent divisions d_i, $i = 2 : n$, were assumed to occur simultaneously. A prosthetic growth curve (bottom) is then generated until the terminal observation time. The calculated growth curve for the missing branch was then added to lineage growth curve

An average growth curve was generated for all lineages within an experimental set. This represents one simulation for the estimation of the global growth curve. Many simulations were performed and were then averaged to create a global growth curve.

The loss of traceable cells also impacted the calculation of the exit kinetics (Fig. 2.6 A). This issue was managed by placing bounds on the expression level of the signal once a cell had been lost (Fig. 2.6 B). The "weak" estimation places an upper bound on the signal by setting the signal to a constant, which is the terminal value observed for that cell. The "strong" estimate creates a lower bound by using using the empirically derived decay function for the signal. The half-life of Rex1-GFPd2 was calculated by least-squares regression analysis for the time-course of Rex1-GFPd2 signal in 2i ES cells and N2B27 cells, post cyclohexamide treatment (Fig. 2.7). The half-life was determined to be different depending upon culture condition. The "strong" estimate was calculated for cells exiting in N2B27, so the half-life from N2B27, $t_{1/2} = 1.46h$, was used. The weighted weak and strong estimates were then used to calculate the mean curves for the exit kinetics.

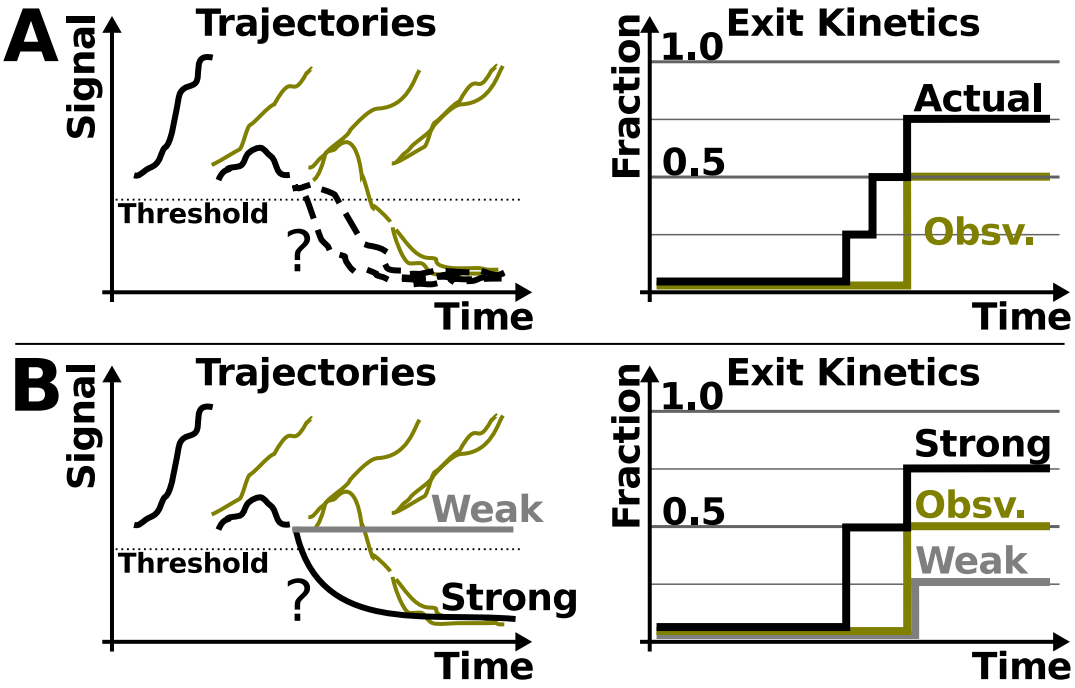


Fig. 2.6 Approximation of Single-Cell Exit Kinetics. (A) When a branch is lost (left, black lines), the exit kinetics that were "observed" (right, gold line) were no longer the "actual" (right, black line) exit kinetics. (B) Two bounding curves have been used to estimate the "actual" exit kinetics. The weak estimate (gray) forms an upper bound by setting that trajectory equal to the terminal observation for all remaining time-points (left). The strong estimate (black) forms a lower bound by assuming that at the moment the cell was lost, the reporter was silenced (left). Thus the trajectory proceeds by exponential decay. When a cell was lost, all subsequent cells on that branch were lost as well. Therefore when the exit kinetics were calculated, a weight was prescribed to the strong and weak estimates which was proportional to the fraction of cells that it would represent in a fully bifurcated lineage-tree. Thus, the fraction of Rex1-GFPd2 negative cells (right) are bounded above by the Strong estimate (black line) and bounded below by the weak estimate (gray line).

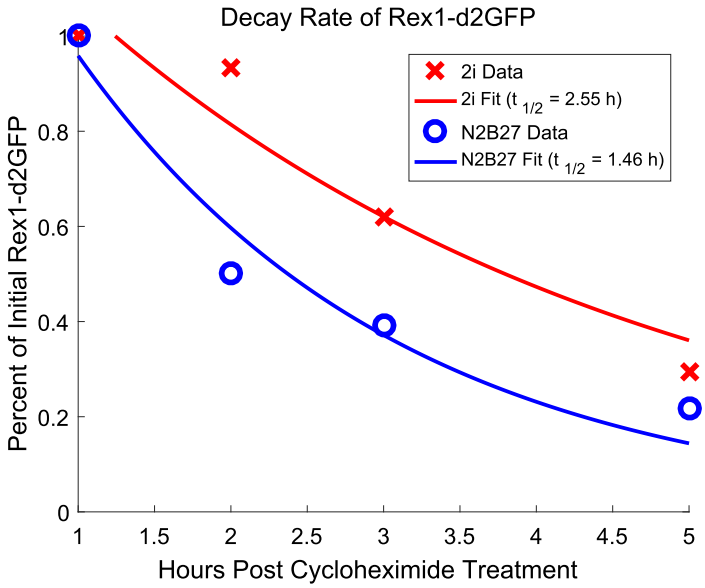


Fig. 2.7 Rex1-GFPd2 Exponential Decay Fitting. The exponential decay function used for the "Strong" estimate were determined by fitting the Rex1-GFPd2 levels after treatment with cyclohexamide (data found in Kalkan et al., 2017). The decay rates were different in 2i (2.55h, $R^2 = 0.926$) and in N2B27 (1.46h, $R^2 = 0.950$). The N2B27 half-life, $t_{1/2} = 1.46h$, was used for the "Strong" estimate because the assays conducted for this portion of the thesis also used N2B27.

The caveat with the strong and weak estimations is that this only holds for the case when a process is either permanently shutting off a signal or permanently turning on a signal. This example uses the case of shutting off a signal permanently. However, it is possible to adapt this for the case of turning on a signal. The difference being that the weak estimation becomes the lower bound and the strong estimate become the upper bound.

Chapter 3

The Behavior of Embryonic Stem Cells Upon Release From Self-Renewal

3.1 Introduction

In this chapter, I study the synchronicity of exit from naïve pluripotency and the relationship between the cell-cycle and exit for ES cells. Experiments were first conducted using the 2i system. This provides a homogeneous, naïve starting population (Nichols and Smith, 2009; Wray et al., 2010; Ying et al., 2008). All experiments were then repeated in a serum-based medium to allow for comparisons to traditional serum/LIF cultures. However, the serum/LIF culture system yields a heterogeneous starting population (Chambers et al., 2007; Hayashi et al., 2008; Toyooka et al., 2008), which has the potential to confound the synchronicity of the exit process. Thus, the Mek inhibitor PD03 was added to serum/LIF. This provided a more homogeneous starting population (Wray et al., 2010; Ying et al., 2008).

In the first set of assays, I set out to examine the synchronicity of exit from the naïve state. Current work has noted that down-regulation of the core pluripotency circuitry occurs asynchronously upon withdrawal of self-renewal signals (Kalkan et al., 2017). However, these kinetics have not been fully quantified. During this phase of down-regulation, there exists a window of time in which cells are able to be recovered back to naïve pluripotency (Betschinger et al., 2013; Kalkan et al., 2017; Kalkan and Smith, 2014; Leeb et al., 2014; Martello and Smith, 2014; Mulas et al., 2017). This is accomplished by reapplication of self-renewal signals. However, after this window of time closes, cells that no longer remain in this reversible state and are said to have exited. In a complementary study to that from Kalkan et al., 2017, I quantify the recoverability of ES cells as they exit naïve pluripotency.

The cell-cycle is inherently linked to ES cells by self-renewal, but does it have a role in differentiation? The state of cell-cycle has been implicated in lineage specification. In human ES cells, it is thought that cyclins regulate the pro-endodermal differentiation signaling of the Tgf-β-Smad2/3 pathway (Pauklin and Vallier, 2013, 2014). It has recently been suggested that mouse ES cells can only exit pluripotency upon the completion of the cell-cycle (Waisman et al., 2017), however, it is difficult to dissect cause and effect when disrupting something as critical as the cell-cycle. Other studies have seen different doubling times for serum/Lif and 2i (Kolodziejczyk et al., 2015) and cycling times that differ between the subpopulations within the heterogeneous serum/Lif cultures (Herberg et al., 2016). So, for the second set of assays, I have interrogated the cell-cycle and its relationship to the exit process. I compared the doubling times of ES cells in self-renewal conditions and of exiting ES cells to monitor any differences during the initial step away from the naïve state.

3.2 Exit from the ES Cell State is Asynchronous

Time-course recovery assays were conducted to quantify the synchronicity of recoverability at both the population-level and clonal-level. Cells cultured in 2i on laminin in 5% O_2 were subjected to withdrawal conditions for set time-periods. Self-renewal signals were reapplied at the end of these time-periods and cells were given the opportunity to respond. If the cells have exited, they will not respond, but if they are still dwelling in a reversible state, they will be recovered back to the naïve state (Betschinger et al., 2013; Kalkan et al., 2017; Leeb et al., 2014).

This assay was conducted at two different densities with corresponding quantification methods. The first condition was a higher density population-level assay for which I used flow cytometry to measure the expression levels of Rex1-GFPd2. The second assay was performed at clonal density. This assay was quantified by staining, imaging, and qualitative scoring of individual colonies. Rex1-GFPd2 expression and a second naïve pluripotency marker, alkaline phosphatase (AP) (Kalkan et al., 2017), were used to score colonies as positive, negative, or mixed.

The kinetics of exit define the cumulative distribution function (CDF) with time as the random variable. From this distribution, I was able to quantify the mean and standard deviation of the time until exit. These data allowed me to determine whether ES cells are exiting together, synchronously, or at different times, asynchronously. Exit from the naïve state is irreversible without reprogramming (Kalkan et al., 2017), therefore the fraction of recoverable cells should be non-increasing in time. If exit is a synchronous process, one

would expect to see a step function, where all cells remain in the naïve state up to a certain time and then all cells exit. Alternatively, if this process is asynchronous, one would expect to see a function that gradually decreases in the proportion of cells that are recoverable (e.g. exponential, sigmoidal, etc.).

3.2.1 Population-Level Recovery Kinetics

ES cells cultured in 2i were subjected to self-renewal signal withdrawal by a media change from 2i to N2B27, the basal medium, for different periods of time (Fig. 3.1). Chiron and PD03 were then reapplied to the cells, which were then permitted time to respond. The cultures were then analyzed by flow cytometry for expression of Rex1-GFPd2.

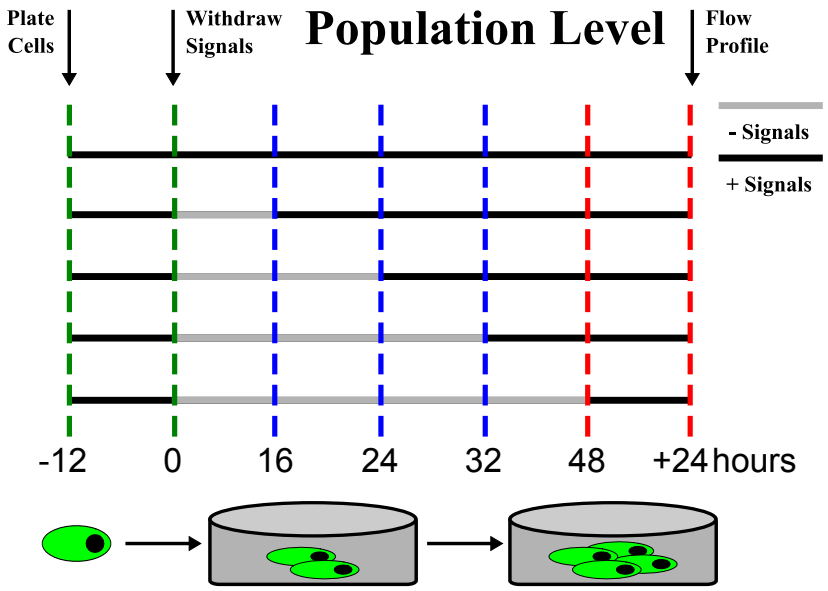


Fig. 3.1 Population-Level Exit Kinetics Assay. Rex1-GFPd2 cells were plated on laminin at a density of 5000 cells/10 cm^2 in 2i. After 12 hours, the medium was aspirated and replaced with N2B27. 2i was serially reapplied for the different time-points: 16, 24, 32, 40, and 48 hours post withdrawal. A positive control was maintained in 2i throughout the withdrawal time-course. Cells were allowed to recover in 2i for at least 24 hours, after which flow cytometry was performed on all samples.

Flow profiles for all biological replicates are shown in Fig. 3.2. A bimodal distribution of the naïve Rex1-GFPd2 positive cells and exited Rex1-GFPd2 negative cells can be seen. These two populations were quantified by applying a systematic manual threshold (Fig. 3.3). Profiles were first gated by forward scatter area and side scatter area. Gated cells were then plotted on a $\log_{10}$ scale for FITC (green) and PE (red). The positive control and the final time-point were used to define the gates for the Rex1-GFPd2 positive and negative

populations, respectively. These gates were then applied to all other time-points within the biological replicate. Percentages of the two populations were then quantified by this gating.

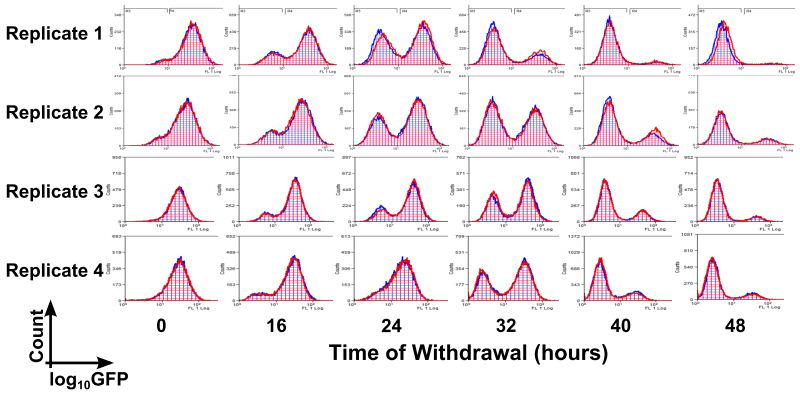


Fig. 3.2 Population-Level Exit Kinetics Flow Profile Time-Course. The time-course of flow profiles for the population-level recovery assay are shown from left to right. Each flow profile shows the log_{10}GFP on the x-axis and the count on the y-axis. Four biological replicates were performed in technical duplicate, shown in red and blue. Each profile is comprised of at least 10,000 cells. (n = 8). A bimodal distribution of Rex1-GFPd2 positive and negative cells appear, with fewer Rex1-GFPd2 cells the longer cells remain without 2i.

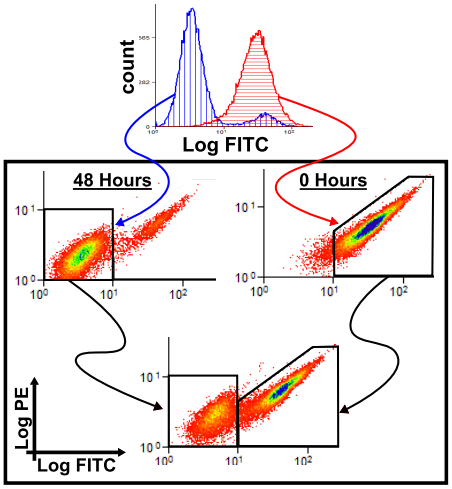


Fig. 3.3 Quantification of Flow Cytometry. Gating for positive and negative populations was performed using the positive control (red histogram) and final time-point (blue histogram), respectively. Black Box: Gates were defined on the two dimensional FITC v PE histogram. The Rex1-GFPd2 negative gate was made by construing a box around the population cloud in the lower-left corner of the 48 hour sample (top left). The Rx1-GFPd2 positive gate was made by constructing a non-overlapping polygon around the second distinct cloud above 1000 AU FITC (top right). These gates were then used for all other time-points within the same biological replicate (Example gating shown at the bottom).

The mean of each biological replicate can be seen in Fig. 3.4. Each replicate generates a sigmoidal CDF, with time as the random variable. The onset of exit begins no later than 16 hours after self-renewal signal withdrawal, with the most dynamic time being between 24 and 32 hours. By 48 hours, nearly the entire population is Rex1-GFPd2 negative.

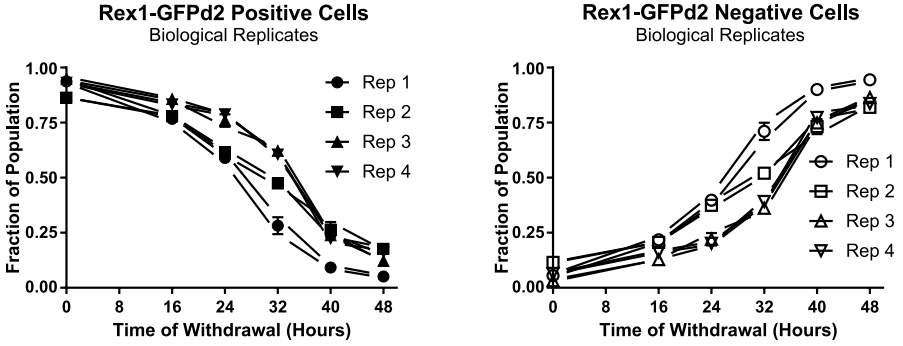


Fig. 3.4 Population-Level Exit Kinetics Biological Replicates. Shown above are replicates from all biological replicates of the population-level withdrawal assay. Each data-point is the mean $\pm$ SEM of the technical replicates. Each line connecting data-points is the individual technical replicate. Left, percent of Rex1-GFPd2 positive cells at time t. Right, percent of Rex1-GFPd2 negative cells at time t. (n = 8)

The Weibull distribution was fit to the data and used to calculate the mean and standard deviation of time to exit (Eqn. 3.1).

$$\mathbf{CDF}(t;\lambda,\gamma) = \begin{cases} 1-e^{-(t/\lambda)^{\gamma}} & , t \geq 0 \\ 0 & , t < 0 \end{cases} \tag{3.1}$$

The Weibull distribution has been selected for several reasons. Foremost, the random variable is time, which can only take on real, non-negative values. Also, when these data are averaged a sigmoidal curve takes form (Fig. 3.5), which is the shape of the Weibull CDF for $\gamma > 1$.

At a higher level, the exit from pluripotency can be thought of as the collapse, or failure, of a transcription factor network (Dunn et al., 2014; Ng and Surani, 2011; Young, 2011). The Weibull distribution was first derived by Fréchet, 1927, and first applied to a physical system by Rosin, 1933, to describe the distribution of particle size in coal dust. It was latter used by Weibull et al., 1951, to model the failure of industrial materials and processes. Since then, it has been derived from first principles for other industrial and biological problems, prescribing physical meaning to the parameter space (Brown and Wohletz, 1995; MacArthur, 1957; Stauffer, 1979).

Population exit times, when fit to the Weibull distribution, provide a mean exit time of $\bar{t} = 31.4$ hours with a standard deviation of $SD = 13.6$ hours. The kinetics proceed by an

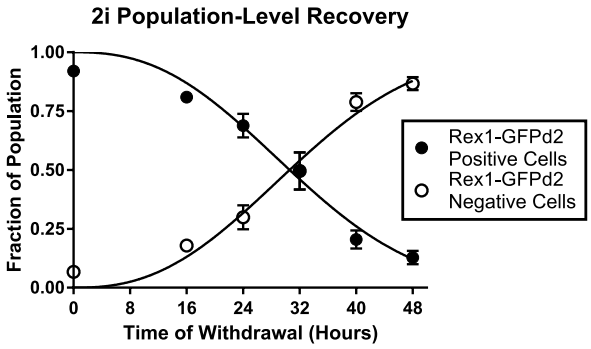


Fig. 3.5 Mean Population-Level Exit Kinetics. The mean ± SEM of all replicates from population-level withdrawal assay. Closed circles are Rex1-GFPd2 positive cells, open circles are Rex1-GFPd2 negative cells. The Black lines were fitted to the data using the Weibull distribution. The mean time of exit is $\bar{t} = 31.4$ with a standard deviation of $SD = 13.6$. The equation parameters are $\lambda = 35.5$ $\gamma = 2.5$ with an $R^2 = 0.97822$.

initial latency period where cells do not exit. Once the process begins, the majority of cells exit within about 30 hours, as determined by the standard deviation. This is on the same scale of exit *in vivo*. In the embryo, the appearance of naïve pluripotency to the dissolution of the naïve circutry occurs happen within one day.

3.2.2 Clonal Recovery Kinetics

The same assay performed in Sec. 3.2.1 was repeated at clonal density. This was done to determine if the exit kinetics and timing of exit are density dependent. Additionally, clonal density offers the benefit of providing smaller, more homogeneous colonies, further reducing the chance of introducing confounding cellular heterogeneity. Results were then analyzed via imaging and qualitative colony scoring (Fig. 3.6).

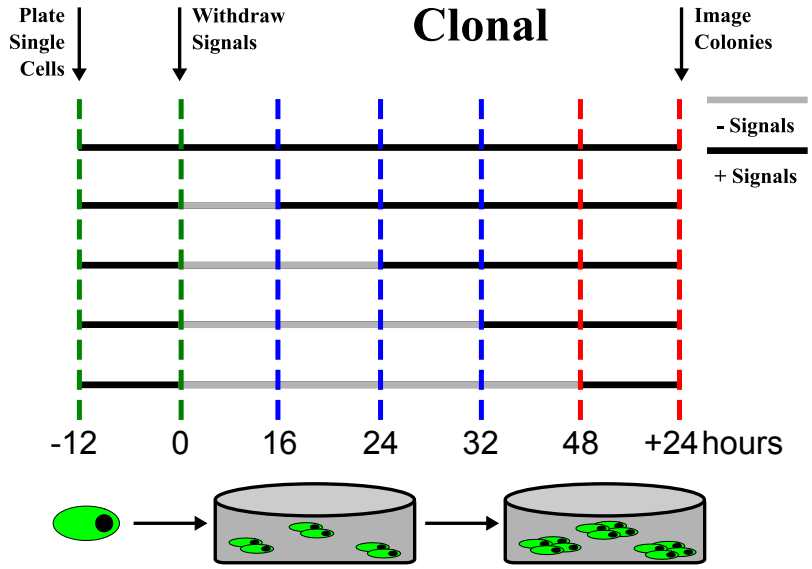


Fig. 3.6 Clonal Exit Kinetics Assay. Rex1-GFPd2 cells were plated on laminin at single-cell density (500 cells/10 cm^2) in 2i. After 12 hours, the medium was aspirated and replaced with N2B27. 2i was serially reapplied for the different time-points: 16, 24, 32, 40, and 48 hours post withdrawal. A positive control was maintained in 2i throughout the withdrawal time-course. Cells were then allowed to recover in 2i for at least 24 hours, at which point each sample was stained for nuclei (Hoechst), alkaline phosphatase (AP), and Rex1-GFPd2 (GFP) and imaged.

Colonies were qualitatively scored as being positive, negative, or mixed for the AP and GFP markers, as seen in Fig. 3.7. A true positive colony will have a compact morphology with ubiquitous expression of AP and GFP, negative colonies will have a more diffuse morphology with no AP or GFP expression, and a mixed colony will have a diffuse morphology with interspersed islands of AP and GFP expression. With this in mind, looking at the raw data in Fig. 3.8, we see fewer AP and GFP positive colonies are able to be recovered the longer the cells are held in withdrawal conditions.

Fig. 3.9 and Fig. 3.10 show the quantification of scoring for GFP and AP colonies, respectively. Each figure shows the percentage of positive, negative, and mixed colonies relative to the total number of colonies present. A similar sigmoidal curve is seen thus indicating an asynchronous exit. The mean exit time is similar at about 32 hours, however, the data is more spread, with a standard deviation of 20 hours, as compared with the 13.6 hours seen in the population data.

To establish a reason for the difference in the spread of the exit times, I considered the total number of colonies generated from each time-point (Fig. 3.11). The data in this assay was considered on a percentage basis to account for variability in the number of colonies

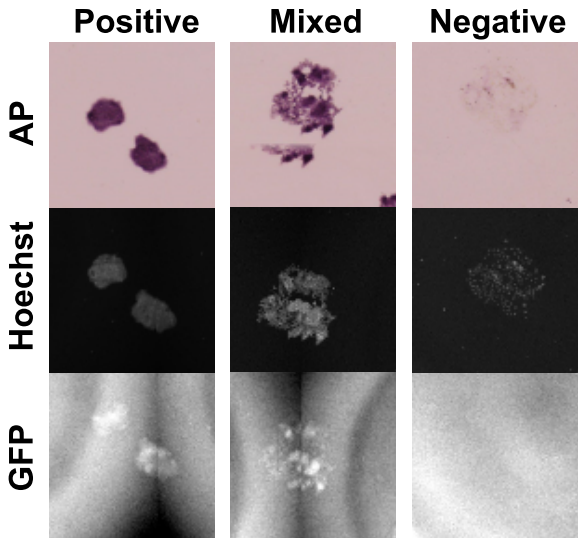


Fig. 3.7 Colony Types for Qualitative Scoring. The clonal assay was analyzed by qualitative scoring of the colony by alkaline phosphatase staining (AP), top row; Rex1-GFP1d2 fluorescence, bottom row; and morphology. The middle row shows nuclei staining by Hoechst33342. The columns show representatives of the three types of colonies scored. Two "positive" colonies can be seen in the left column, which are compact and both AP and GFP positive. The "mixed" colony in the center column is larger with interspersed AP and GFP signal. The "negative" colony is sparce with neither AP nor GFP expression. These images are cropped directly from Fig. 3.8 time-points 0, 24, and 32, respectively.

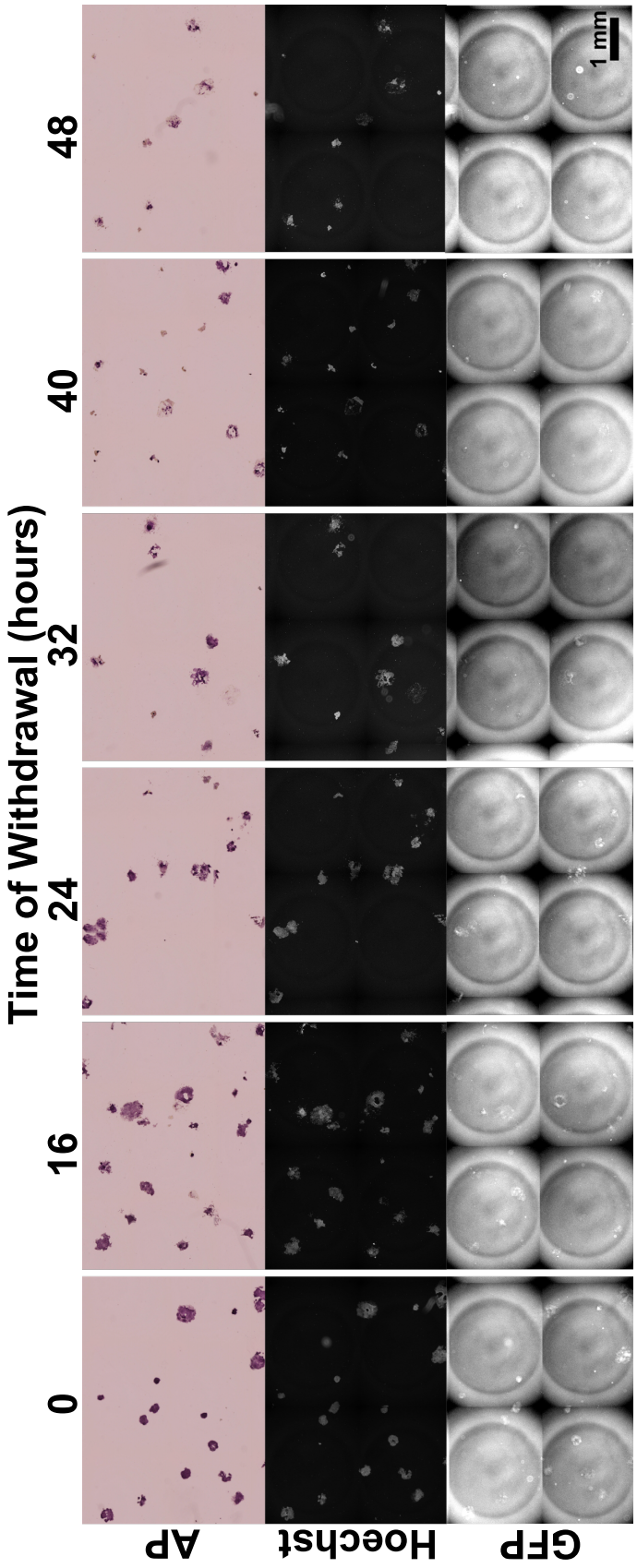


Fig. 3.8 Clonal Exit Kinetics Image Time-Course. Above are representative images from the clonal exit kinetics time-points. Each row shows alkaline phosphatase (AP) staining, nuclear (Hoechst) staining, and Rex1-GFPd2 (GFP) fluorescence for the time-course of recovery. The longer cells remain without self-renewal signals, the fewer ES cell colonies (AP and GFP positive) can be recovered.

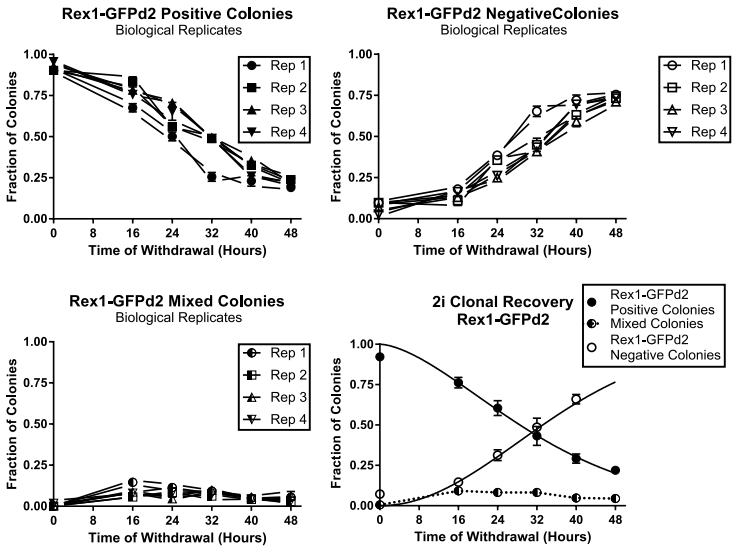


Fig. 3.9 Clonal Exit Kinetics Using Rex1-GFPd2. Biological replicates for Rex1-GFPd2 clonal recovery. Each data-point is the mean ± SEM of the technical replicates and each line is the individual technical replicate. **Top:** Left, percent of Rex1-GFPd2 positive colonies at time t. Right, percent of Rex1-GFPd2 negative colonies at time t. **Bottom:** Left, percent of Rex1-GFPd2 mixed at time t. Right, the mean ± SEM of all replicates. The mean time to exit is $\bar{t} = 32.3$ with a standard deviation of $SD = 02.5$ ($R^2 = 0.9824$, $\lambda = 36.0$, $\gamma = 1.6$). (n=8)

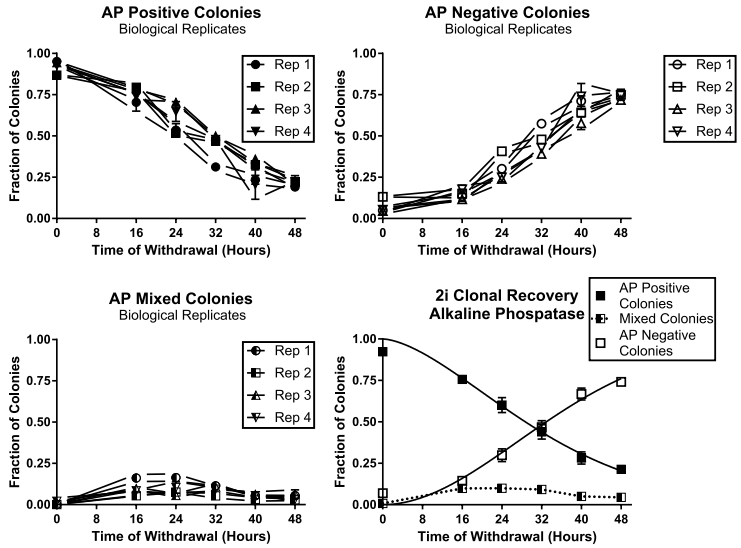


Fig. 3.10 Clonal Exit Kinetics Using Alkaline Phosphatase. Biological replicates for AP clonal recovery. Each data-point is the mean ± SEM of the technical replicates and each line is the individual technical replicate. **Top:** Left, percent of AP positive colonies at time t. Right, percent of AP negative colonies at time t. **Bottom:** Left, percent of AP mixed at time t. Right, the mean ± SEM of all replicates. The mean time to exit is $\bar{t} = 32.1$ with a standard deviation of $SD = 20.3$ ($R^2 = 0.9823$, $\lambda = 35.8$, $\gamma = 1.6$). (n=8)

formed across replicates. However, a direct comparison across all data only holds if N is consistent. Fig. 3.11 shows that the number of colonies formed in the 0 and 16 hour time-points are statistically greater than all other time-points. This information is summarized in Table 3.1.

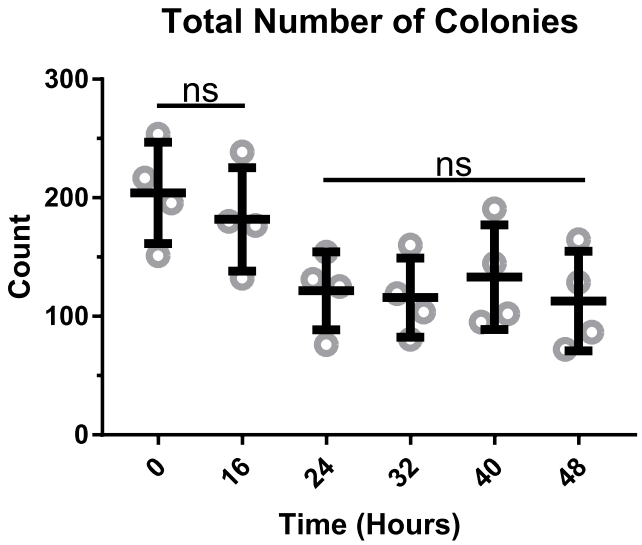


Fig. 3.11 Number of Colonies from Clonal Exit Kinetics Biological Replicates. Fewer colonies were recovered from later time-points, indicating an increase in cell death.. The gray circles are the total number of colonies for each biological replicate and the black bars are the mean and standard deviation. Time-points 0 and 16 are statistically significant to time-points 24, 32, 40, and 48 (p-value < 0.0001). (n=8)

Table 3.1 Mean Number of Colonies from Clonal Exit Kinetics. Mean number of colonies assessed for each technical replicate in the clonal recovery assay corresponding to Fig 3.11. A significant decrease is seen between the 16 and 24 hour time-points. (n=8)

Number of Colonies Assessed per Biological Replicate			
Time-point (hours)	**Mean**	**SD**	**N**
0	204.1	42.2	4
16	181.8	43.6	4
24	121.5	32.8	4
32	115.8	33.5	4
40	133.0	44.1	4
48	112.9	41.9	4

It is plausible that the reduced number of colonies in the later time-points have skewed the data. Culturing cells at clonal density tends to be a more stressful environment. Additionally, the application of 2i to non-naïve cells can lead to apoptosis (Nichols and Smith, 2009). These, and other, caveats could result in reduced viability for cells that have exited in the later time-points. That would imply that the fraction of exited colonies should be greater, and, if they were considered, the resulting kinetics would yield a less spread distribution. Thus, I suggest, as one possibility, that an underestimation of exited colonies has lead to the difference in the exit kinetics on the clonal scale.

3.2.3 Summary

Using two independent methods, I have demonstrated that exit from naïve pluripotency is an asynchronous process which occurs over a two day period. The evidence comes by way of the sigmoidal exit kinetics curves for both the population-level recovery and clonal recovery assays. These data are summarized in Table 3.6. Had this been a synchronous process, a step function would have been expected, which is not the case. From these data I have quantified the distribution of exit times in both conditions.

Table 3.2 Summary of Exit Kinetics in 2i. Quantification of Exit Kinetics from 2i in N2B27 for the different methods of analysis. Shown are the mean time to exit ($\bar{t}$), the standard deviation (SD), the scale (λ) and shape (γ) parameters of the Weibull distribution, and the goodness of fit (R^2).

Conditions	$\bar{t}$	**SD**	λ	γ	$\mathbf{R}^2$
Population	31.4	13.6	35.5	2.5	0.9782
GFP	32.3	20.5	36.0	1.6	0.9842
AP	32.1	20.3	35.8	1.6	0.9823

It was observed that the two assays had slightly different dispersions. A plausible explanation for the difference of the standard deviation between the two assays was provided by showing that in the clonal recovery assay, there was an underestimation in the total number of colonies recovered in the later time-points. Otherwise the conclusion is that higher density results in a more uniform exit compared to low densities.

3.3 Doubling Times are Maintained During Exit from the ES Cell State in Chemically Defined Medium

It has been reported that doubling times vary in different culture conditions (Kolodziejczyk et al., 2015) and within heterogeneous serum/LIF cultures (Herberg et al., 2016). Therefore, I compared the doubling times of ES cells in self-renewal conditions and ES cells differentiating to determine if any change occurs during the initial step away from the naïve state. Doubling times for both cases were determined by performing a similar assay to Sec. 3.2.2 (Fig. 3.12). The difference here is that samples were stained with Hoechst and imaged at high resolution at the end of each time-point.

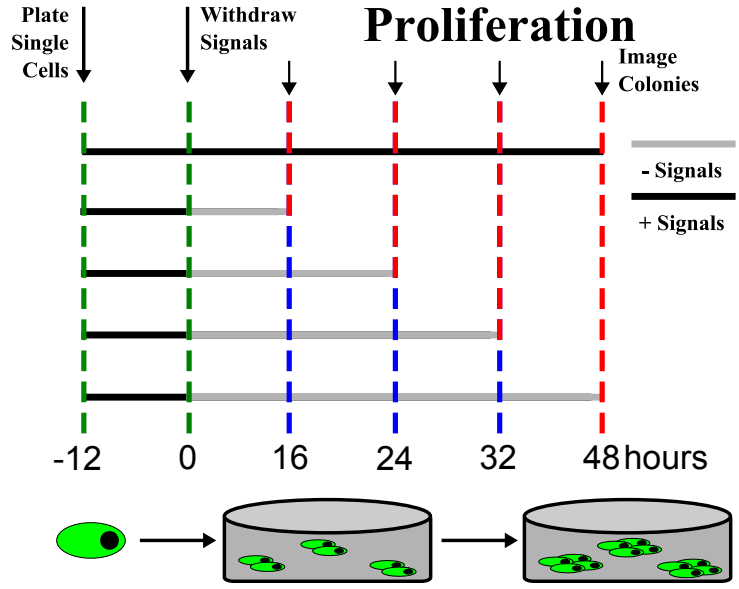


Fig. 3.12 Proliferation Kinetics Assay. Rex1-GFPd2 cells were plated on laminin at single-cell density (500 cells/10 cm^2) in 2i. After 12 hours, the medium was aspirated and replaced with either 2i or N2B27. At each time-point (16, 24, 32, 40, and 48 hours) samples from the 2i and N2B27 conditions were stained for nuclei (Hoechst) and imaged at high resolution.

3.3.1 Quantification of Doubling Time

To assess doubling times, distributions of the number of cells per colony were generated for each time-point (Fig. 3.13). Samples were then stained with Hoechst and imaged at a sufficient resolution to resolve individual nuclei. Randomly selected nuclei were manually segmented and quantified for fluorescence intensity after background correction. These values were then used to generate a linear calibration curve of fluorescence intensity vs. number of

nuclei (1, 5, 10, and 20 cells). This was performed for each time-point, selecting nuclei from across technical replicates, to account for variability across images. Once calibration curves were generated, the total fluorescence of approximately 100 colonies, from each technical replicate, was quantified. The number of nuclei per colony was then calculated from the regression equation. Linear regression analysis results from each biological replicate can be seen in Table 3.3. All calibration curves generated for determining the number of cells per colony were consistent with high R^2 values.

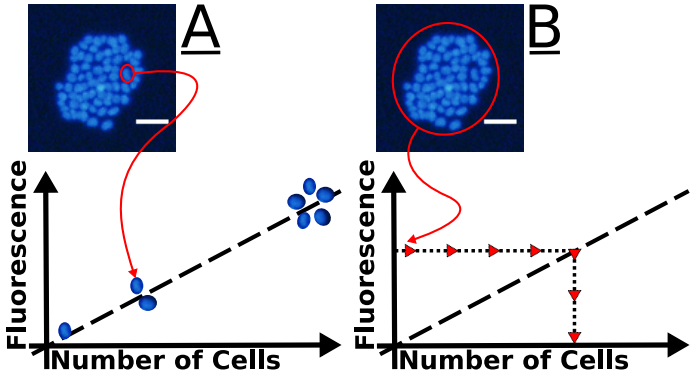


Fig. 3.13 Quantification of Proliferation Kinetics **A.** Once images were taken, calibration curves were generated from fluorescent images by manually segmenting randomly selected nuclei, which were stained with Hoechst (cal. curve 1, 5, 10, and 20 nuclei). Linear regression was performed on the total intensity for the different number of nuclei to obtain the equation of the line. Calibration curves were generated for each plate by taking an equal number of nuclei from each well. **B.** To determine the number of nuclei per colony, the total fluorescence of the colony was acquired by manual segmentation. The resulting value was background-corrected by the mean intensity of a local region without cells. This fluorescence value was then used to calculate the nuclei within that colony. (scale bar = 50 μm)

3.3.2 Proliferation Kinetics

Cells were cultured in either 2i or N2B27 on laminin in 5% O_2 at clonal density. Cell nuclei were then stained with Hoechst and imaged at high resolution at set time-points. Calibration curves for fluorescence per nuclei were generated for each time-point. These calibration curves were then used to determine colony sizes from total fluorescence of a colony. Representative images of the time-course can be seen in Fig. 3.14.

Technical replicates were pooled together to generate a large enough N for statistical purposes (Table 3.4). Between 300 and 400 colonies were assessed per biological replicate, resulting in between 1,300 and 1,500 colonies for each time-point. As time progresses, colony sizes become larger, as can be seen in Fig. 3.15.

Table 3.3 Regression Analysis for Proliferation Kinetics Quantification. Shown below are the slope, y-intercept, and R^2 value for the equations generated by linear regression analysis. Each calibration curve is shown for each time-point from each biological replicate. The same equation was used to quantify both the 2i and N2B27 samples, being that nuclei from both samples were used to generate the calibration cure at each time-point. High levels of precisions are shown to be reproducible as indicated by the goodness-of-fit R^2-value.

time-point (hours)	slope	y-intercept	R^2
16	3484	767.5	0.9998
	3320	720.7	0.9990
	3242	-858.8	0.9989
	3301	353.8	0.9985
24	3320	427.4	0.9988
	3298	379.1	0.9997
	3435	191.8	0.9973
	3303	-737.4	0.9995
32	3410	-249.3	0.9946
	3052	205.8	0.9997
	3591	-939.6	0.9996
	3643	-205.0	0.9947
40	3884	-207.5	0.9999
	3247	-496.1	0.9997
	3142	198.8	0.9982
	3303	438.8	0.9996
48	4031	-337.8	0.9994
	3097	-188.8	0.9996
	3946	350.5	0.9884
	3422	-676.6	0.9992

Table 3.4 Proliferation Kinetics Mean Number of Colonies Assessed. The mean, stand deviation and number of biological replicates are shown for the number of colonies assessed per biological replicate. Over three hundred colonies were assessed for each biological replicate.

Number of Colonies Assessed per Biological Replicate						
	2i			**N2B27**		
time-point (hours)	**Mean**	**SD**	**N**	**Mean**	**SD**	**N**
16	335.75	42.21	4	361.00	41.09	4
24	357.25	61.26	4	346.25	43.60	4
32	369.25	30.31	4	371.75	23.88	4
40	380.00	50.94	4	377.25	64.92	4
48	342.75	48.97	4	338.00	50.94	4

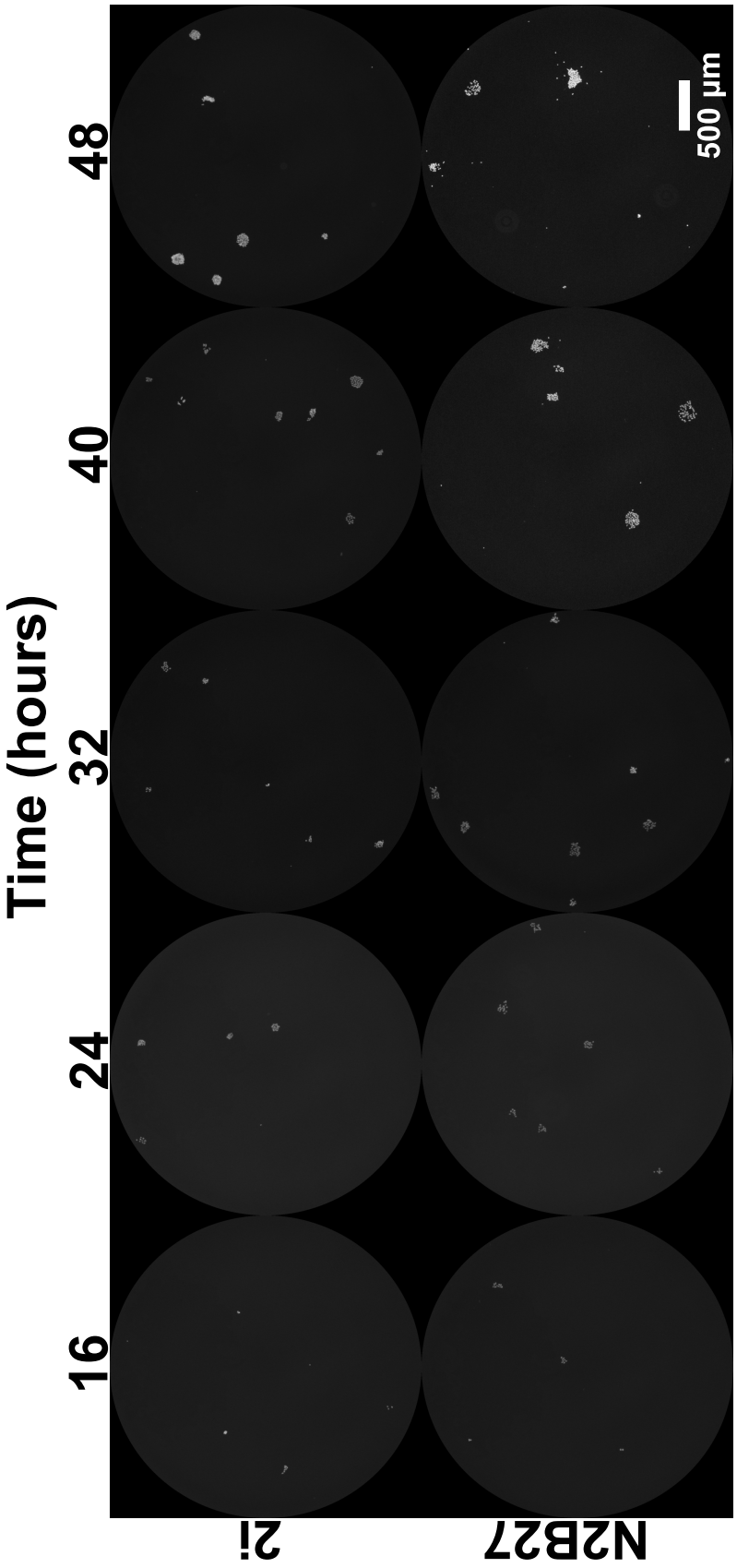


Fig. 3.14 Proliferation Kinetics Image Time-Course. Representative snapshots of Hoechst staining from the proliferation kinetics assay. Each row shows the time-course of colony formation in either ES cell supporting 2i (top row) or pro-differentiation conditions N2B27 (bottom row). As time progress individual colony size increases in both conditions. Similar to Fig. 3.8, the larger 2i colonies show a more compact morphology as compared to the more diffuse N2B27 colonies.

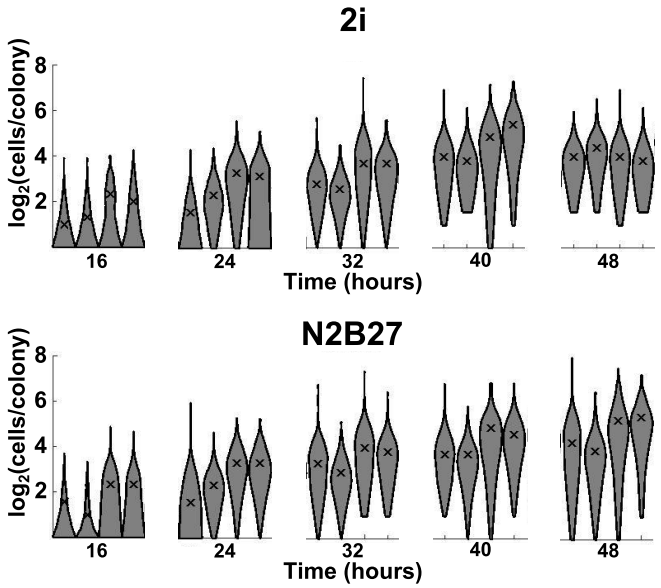


Fig. 3.15 Proliferation Kinetics Violin Plots. The Violin plots of biological replicates for cells cultured in 2i (top) and N2B27 (bottom). Colony size for each time point was determined by taking mean of all colonies assessed and these values were used to generate growth curves for each biological replicate. The 'x' indicates the median.

Growth curves for each biological replicate were generated (Fig. 3.16) and transformed to a semi-$\log_2$ scale. Doubling times were calculated for each biological replicate for both conditions. This was accomplished by fitting a linear curve to the linearized growth data, finding the slope of that curve, and using the formula,

$$\text{doubling time} = \log(2)/slope. \tag{3.2}$$

These results are summarized in Table 3.5. The mean doubling time is the mean of the doubling times from the four biological replicates. A batch effect can be seen (Fig. 3.16) between the first two biological replicates and the last two biological replicates. However, this is not an issue, because the slope of the linearized growth curve is the relevant information, and this remains unaffected by the batch effect. t-Test shows no statistical significance between the two conditions (p-value = 0.9920), indicating that the doubling time is not affected by exit from naïve pluripotency.

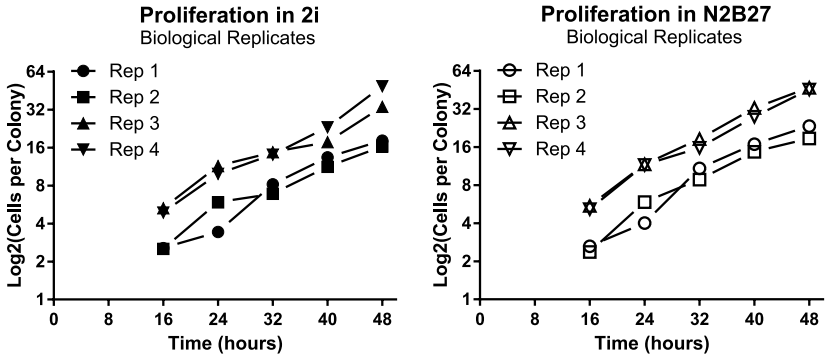


Fig. 3.16 Proliferation Kinetics Biological Replicates Growth Curves. The mean colony size from each time-point was used to growth curves for 2i (left) and N2B27 (right). Each data-point is the mean ± SEM. Each curve was used to calculate the doubling time for the biological replicate. Time is shown on the x-axis and the $\log_2$ number of cells per colony is shown on the y-axis. ($n = 4$)

Table 3.5 Doubling Times in Chemically Defined Cultures. Doubling times for cells cultured in 2i and N2B27 were calculated from the growth curves in Fig. 3.16. No significant difference is seen between the doubling times of ES cells in 2i and differentiating cells in N2B27. For each replicate, the doubling is shown and the final last row ('Mean') shows the mean ± SEM of the biological replicates.

Doubling Time (hours)		
	2i	**N2B27**
Replicate 1	12.36	11.92
Replicate 2	13.82	13.77
Replicate 3	13.16	11.83
Replicate 4	9.11	10.98
Mean	12.11 ± 1.044	12.13 ± 0.5878

3.3.3 Summary

I set out to determine if there was a difference in the doubling time between 2i ES cells and N2B27 cells. This was accomplished by the collection of colony size distribution over a time-course via high resolution imaging and fluorescence quantification. No difference was seen between the two conditions (p-value = 0.9920), with 2i ES cells doubling every 12.11 hours and N2B27 cells doubling every 12.13 hours.

3.4 Serum-Based Cultures Exit on a Larger Time-Scale with a Faster Doubling Time

All assays were repeated in a serum-based medium to draw comparisons with the traditional heterogeneous serum/LIF ES cell cultures. However I am examining the synchronicity of events and serum/Lif begins as a heterogeneous population, which is undesirable for this assay. Any combination of at least two of the three signals, LIF, PD03, and Chiron, are sufficient to yield a naïve, homogeneous ES cell culture (Dunn et al., 2014; Wray et al., 2010), therefore PDO3 was added to generate a homogeneous starting population. All assays in this section are conducted with LIF & PDO3 in Serum medium (LPS) on gelatin at 20% O_2.

3.4.1 Exit from the ES Cell State in Serum-Based Medium is Asynchronous

The same withdrawal assays from Sec. 3.2 were repeated under LPS conditions. The main difference is that the kinetics were approximately twice as long. First, consider the population-level recovery (Fig. 3.17). In LPS, cells do not begin the exit process until after 24 hours. The most dynamic period exists between 48 hours to 96 hours, with most cells exiting by 96 hours. Again, the kinetics take on a sigmoidal shape, indicating the asynchronous nature of the exit process. Fitting the Weibull distribution yields a mean exit time of 71.8 hours with a standard deviation of 23.8 hours.

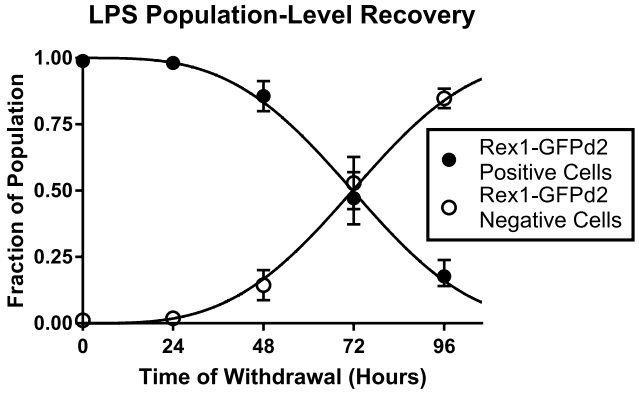


Fig. 3.17 LPS Population-Level Exit Kinetics. The mean ± SEM of all replicates from population-level withdrawal assay. The closed black circles are Rex1-GFPd2 positive cells, the open white circles are Rex1-GFPd2 negative cells. The black lines are fit using the Weibull distribution. This yields a mean exit time of $\bar{t} = 71.8$ with a standard deviation of $SD = 23.8$. The equation parameters are $\lambda = 80.1$ $\gamma = 3.3$ with an $R^2 = 0.9971$.

The clonal recovery kinetics show similar results. The major difference here is the substantial proportion of mixed colonies, which have skewed the positive and negative curves, underestimating what is true positive and true negative, therefore each curve shall be considered separately. The reduction in true positive colonies begins at 24 hours. The most dynamic region exists between 24 and 72 hours with no positive colonies existing at 96 hours. The first appearance of true negative colonies begins after 24 hours, with the most dynamic region exists between 48 and 96 hours. The total elimination of mixed colonies from the culture could not be obtained because the culture was becoming confluent and thus impacting the viability. Although there is still a large proportion of mixed colonies at 72 and 96 hours, the proportion of positive cells in the mixed colonies decreases as the time of withdrawal increases. To obtain kinetics that resemble the population data, individual colonies would need to be assessed for the proportions that are naïve and exited.

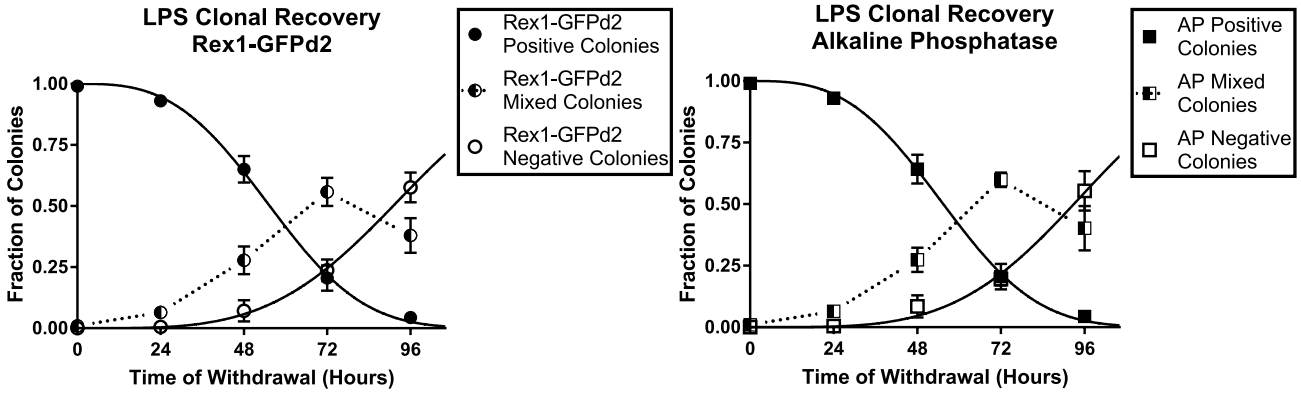


Fig. 3.18 LPS Clonal Exit Kinetics. Left: Biological replicates for Rex1-GFPd2 fluorescence. The mean time to exit is $\bar{t} = 55.9$ with a standard deviation of $SD = 23.8$ ($R^2 = 0.9987$, $\lambda = 62.6$, $\gamma = 3.0$). Right: Biological replicates for alkaline phosphatase staining. The mean time to exit is $\bar{t} = 55.7$ with a standard deviation of $SD = 20.5$ ($R^2 = 0.9990$, $\lambda = 62.4$, $\gamma = 3.0$). (Each data-point is the mean $\pm$ SEM of the technical replicates and each line is the individual technical replicate (n=6). Marker key: Filled = Positive, Half-filled = Mixed, Open = Negative.)

Even with this limitation in clonal assay, the kinetics take on a sigmoidal shape, indicating an asynchronous exit process. The Weibull distribution was fit to the Rex1-GFPd2/AP positive colony kinetics. This was chosen because I am using self-renewal recovery to quantify the proportion of cells that have exited. This yields a mean exit time of 55 hours hours with a standard deviation of of 20 hours.

Table 3.6 Summary of Exit Kinetics in LPS. Quantification of Exit Kinetics from LPS in Serum shown are the mean time to exit ($\bar{t}$), the standard deviation (SD), the scale (λ) and shape (γ) parameters of the Weibull distribution, and the goodness of fit (R^2).

Conditions	$\bar{t}$	**SD**	λ	γ	$\mathbf{R}^2$
Population	71.8	23.8	80.1	3.3	0.9971
GFP	55.9	20.2	62.6	3.0	0.9987
AP	55.7	20.5	62.4	3.0	0.9990

3.4.2 Doubling Time Accelerates During Exit from the ES Cell State in Serum

To determine the doubling time of LPS ES cells and serum cells, the same assay was performed as in Sec. 3.3, with the same time-points as in Sec. 3.4.1. The same quantification method was used, yielding results with the same consistency and precision.

The violin plots in Fig. 3.19 show the increase in the number of cells per colony over time, with some variability between biological replicates. The means of the distributions on the log_2 scale provide near linear growth curves (Fig. 3.20). Regression analysis was performed on the individual replicates and calculated the mean doubling times to be 15 hours in LPS and 10 hours in serum. These data are summarized in Table 3.7. t-Test shows that the doubling time of serum cells is shorter than LPS ES cells (p-value = 0.0096).

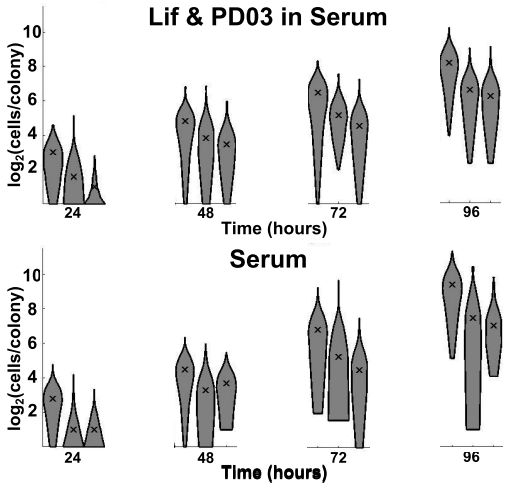


Fig. 3.19 LPS Proliferation Kinetics Violin Plots. Violin plots of biological replicates for cells cultured in LPS (top) and serum (bottom). Colony size for each time point was determined by taking mean of all colonies assessed and these values were used to generate growth curves for each biological replicate. the 'x' indicates the median.

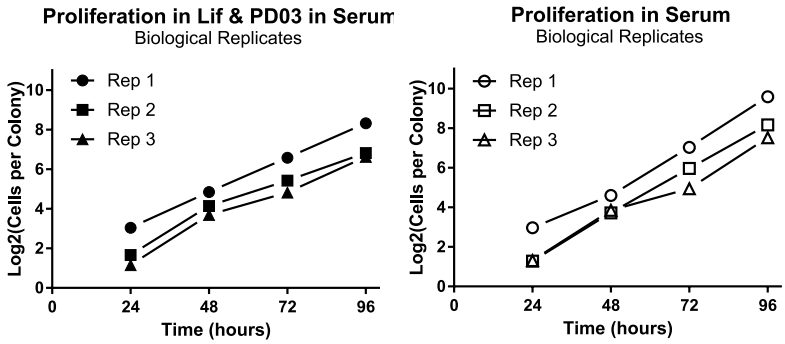


Fig. 3.20 Proliferation Kinetics Biological Replicates Growth Curves. The mean colony size from each time-point was used to growth curves for 2i (left) and N2B27 (right). Each data-point is the mean ± SEM. Each curve was used to calculate the doubling time for the biological replicate. Time of is shown on the *x*-axis and the $\log_2$ number of cells per colony is shown on the *y*-axis. (n = 3)

Table 3.7 Doubling Times in Serum Based Medium. Doubling times for cells cultured in LPS and Serum were calculated from the growth curves in Fig. 3.20. Differentiating cells in Serm had a significantly shorter doubling times than ES cells LPS. For each replicate, the doubling is shown and the final last row ('Mean') shows the mean ± SEM of the biological replicates.

Doubling Time (hours)		
	LPS	**Serum**
Replicate 1	13.75	9.43
Replicate 2	16.77	10.89
Replicate 3	14.25	10.11
Mean	14.92 ± 1.62	10.14 ± 0.73

3.4.3 Summary

The kinetics of exit from the naïve state are sigmoidal regardless of culture condition, density, or method of analysis for the conditions explored. With that, it can be said that the time of exit is asynchronous. The difference being that the kinetics in the LPS condition, compared to 2i, occur on a larger time-scale. An acceleration in the doubling times under withdrawal conditions has also been observed in LPS.

3.5 Failure Analysis for Collapse of the Pluripotency Transcription Factor Network

Upon withdrawal of self-renewal signals, there will be a time at which the pluripotency transcription factor network can no longer be sustained and will "collapse", or fail. The Weibull distribution, Eqn. 3.1, is used to model the failure rate of industrial materials and products. Here it has been fit to the exit kinetics data, drawing the analogy between materials and the transcription factor network.

The parameters of the distribution tell the story of the process in question. The scale parameter λ, units of [time], dictates when failure events take place. The shape parameter γ, unitless, describes how the process fails (Eqn. 3.3). When γ is less than 1, the rate of failure is decreasing in time. When γ is equal to 1, this indicates that the failure is most likely due to random events outside of the process. When γ is greater than 1 the rate of failure is increasing which is indicative of an aging process. To illustrate this concept, consider the rates of death in a population of healthy people over their collective lifetimes. The rate of infant mortality drops as humans get older ($\gamma < 1$), throughout adult life people die from random events like car accidents ($\gamma = 1$), and when people reach old age, their bodies begin to degrade and the rate of death increases ($\gamma > 1$). Changes in the inflection of the CDF due to γ are shown in Fig. 3.21.

$$\gamma \begin{cases} < 1 \Rightarrow \text{decreasing failure rate} \\ = 1 \Rightarrow \text{constant failure rate} \\ > 1 \Rightarrow \text{increasing failure rate} \end{cases} \tag{3.3}$$

In the context of ES cell differentiation, these ideas can be used to find desirable culture conditions for more uniform differentiation. The heterogeneity of *in vitro* differentiation could be mitigated by finding a system in which ES cells rapidly move from pluripotency to the cell type of choice ($\gamma >> 1$ or $\gamma << 1$) in the shortest possible window of time (small λ).

Preliminary data highlights this concept in Fig. 3.21 (right panel). The scale parameter for 2i ES cells is $\lambda = 35.5$. When Chiron is replaced with Lif, the scale parameter doubles to $\lambda = 73.0$ for ES cells in Lif & PD03 in N2B27. A less drastic change is seen when N2B27 is replace with serum, the scale parameter increased to $\lambda = 80.1$. These data are summarized in Table 3.8.

One possible interpretation is that the magnitude of the scale parameter depends upon the stability of the transcription factor network, which is dictated by the extrinsic self-renewal

inputs. If that is the case, than these data would suggest that Lif and PD03 form a more stable network than 2i and that the additional input of Bmp4 from serum further increases this robustness in the LPS conditon.

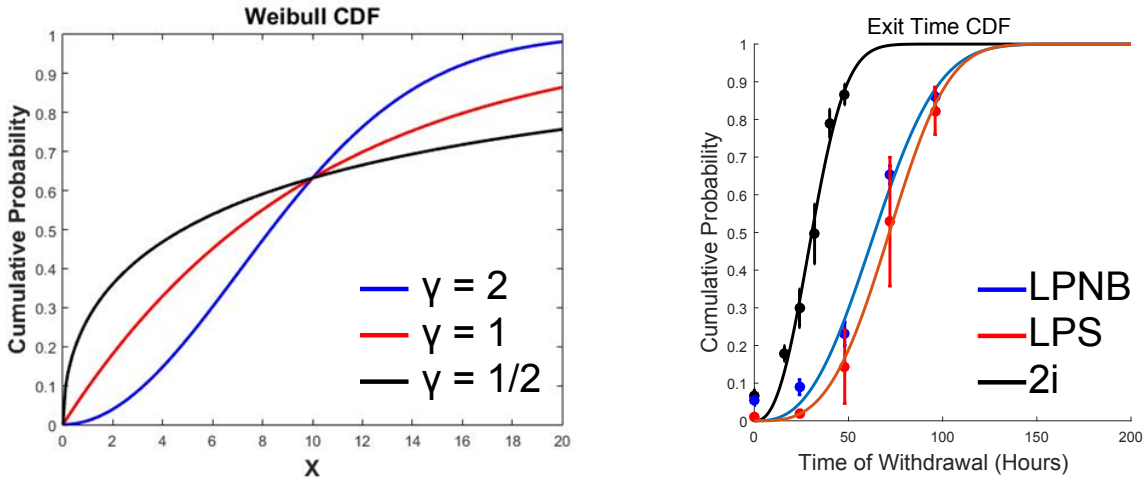


Fig. 3.21 Failure Analysis of the Pluripotency Transcription Factor Network. Left: Effects of shape parameter γ on the Weibull CDF. Increasing γ from less than one to greater than one involves an change of inflection of the curve indicating the type of process that is occurring (see Eqn. 3.3). Right: Empirical fitting of exit kinetics for Lif & PD03 in N2B27 (LPNB), Lif & PD03 in Serum (LPNB), and 2i. All conditions show that that the rate of failure increases over time. However, the the scale parameter λ changes depending upon which of the three self-renewal signals are present. The introduction of Lif shows a lengthening of the exit process.

The shape parameter γ for all three conditions is greater than one, indicating that the pluripotency transcription factor network collapses at an increasing rate as an aging process. The longer that ES cells remain without extrinsic self-renewal signals, the more likely they are to exit naïve pluripotency. This is the case regardless of starting conditions or exit conditions. At the moment these are preliminary data, but these are promising findings to further explore.

Table 3.8 Weibull Distribution Parameter Space for Different Pluripotency Condition. Empirically derived Weibull distribution parameter space for different initial culture conditions. Shown are the three conditions, 2i, Lif & PD03 in N2B27 (LPNB), and Lif & PD03 in Serum (LPS), The scale parameter λ, shape parameter γ, and the goodness of fit R^2. The shape parameter for all three conditions show that exit proceeds with an increasing rate of failure. The conditions using Lif show a larger scale parameter indication that exit takes a longer period of time to progress relative to the condition without Lif.

Conditions	λ	γ	$\mathbf{R}^2$
2i	35.5	2.5	0.9782
LPNB	73.0	2.7	0.9845
LPS	80.0	3.3	0.9782

3.6 Discussion

I have demonstrated using several assays that the exit from naïve state occurs in an asynchronous fashion for both 2i and serum-based cultures. This is in agreement with the current literature (Betschinger et al., 2013; Kalkan et al., 2017; Leeb et al., 2014; Lowell et al., 2006; Zhou et al., 2013). Before cells exit the naïve state, there is a latency period, which varies in duration, where ES cells are able to be recovered back to self-renewal. Furthermore, the distribution of this latency period has now been quantified by considering this process to be a time to failure. In this case, the failure would be the complete collapse of the pluripotency network, as monitored by our proxy marker Rex1-GFPd2. The kinetics curves generated from the recovery assays can be considered to be a CDF with time as the random variable. Fitting this to the failure analysis model, the Weibull distribution, I show that the time to exit in 2i is 32 hours with a standard deviation between 13 and 20 hours. Likewise, in the serum-based culture, the mean exit time is between 56 and 72 hours with a standard deviation of approximately 20 hours, depending upon assay. Further comparison of ES cell exit in Lif & PD03 in N2B27 shows that the main effector on the time-scale difference for 2i and LPS is Lif (this data also informed the long-term single-cell imaging assay, indicating that imaging should take place over at least 48 hours).

The core pluripotency circuitry first emerges in a subpopulation of the nascent ICM in the E3.25 early blastocyst (Boroviak and Nichols, 2014). Over the next day, the ICM will specify into two cell types and self-organize into the extra-embryonic tissue called the primitive endoderm and the naïve epiblast. It is not for another day, the peri-implantation E5.25 embryo, that the core naïve circuitry is extinguished in the epiblast (Boroviak et al.,

2014; Kalkan et al., 2017). The establishment and disassembly of the naïve pluripotency network in the epiblast of the preimplantation mouse embryo occurs over approximately 2 days. *In vitro*, upon withdrawal of self-renewal signals, the 2i culture system requires approximately 2 days for the population to entirely exit, while the LPS system requires 4 days. The goal of capturing *in vivo* cell types *in vitro* not only requires transcripticomic considerations, but it must also consider relevant kinetic behaviors like these. In terms of time scale, the 2i system more closely emulates the life-span of the *in vivo* naïve population as compared to the LPS system. And when taken with the supporting 'omics' data (Boroviak et al., 2014; Kalkan et al., 2017), it builds upon the case in support for the use of the 2i system to study naïve pluripotency.

The doubling times of ES cells and differentiating cells were also considered in both conditions. This yielded doubling times of 12 ± 1 hours for 2i ES cells and 15 ± 1.6 hours for LPS ES cells. Kolodziejczyk et al., 2015, report the doubling times of 25 hours in 2i and 11 hours in serum/LIF. In their study, they perform a standard doubling time assay by plating a known number of cells and then counting the total number of cells accumulated, via hemocytometer, at 24 hour intervals for four days. A direct comparison cannot be made between the serum-based media because they come from different starting population (serum/Lif v LPS), however, the discrepancy of doubling times in 2i can be addressed. A possible explanation for the difference in doubling times in 2i could be that they are using data beyond the exponential phase of growth (Fig. 3.22). They report allowing cultures to grow to near confluency over the four day period over which they perform their proliferation assay. This would be long enough to enter into the stationary phase of growth. They do not show their growth curves, however using data beyond the exponential phase would result in a longer doubling time.

Doubling times were also considered between ES cells and exiting cells. An increased doubling time is observed in serums cells compared to LPS ES cells. The reason for this acceleration of the doubling time as cells initiate differentiation is unclear. There are several possibilities that could alone, or in conjunction, be the cause.

One such case may be linked to behavior seen in the embryo. As ES cells begin the exit from the naïve state, their transcriptome, metabalome, and methylome resemble the E5.5 peri-implantation epiblast, prior to the priming of the post-implantation epiblast (Kalkan et al., 2017). During the formation of the primitive streak, E6.5, "proliferation zones" appear in the epiblast where the rate of proliferation increases to allow for rapid growth (Snow, 1977). If differentiating ES cells continue to progress *in vitro*, as the epiblast does *in vivo*, an onset of rapid proliferation would be expected as these cells begin the process of early

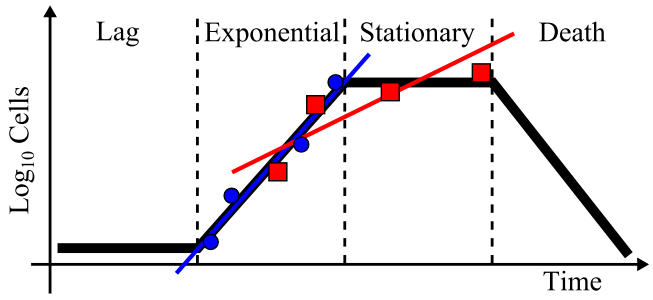


Fig. 3.22 The Phases of Cell Culture Growth. The four phases of cell culture growth. **Points:** data. **Line:** regression line from which doubling times are calculated. **Blue:** Data collected within the exponential growth phase. **Red:** data collected across exponential and stationary phases of growth. To accurately calculate doubling-times, data should be collected within the exponential phase of growth. Collecting data outside of this window will yield inaccurate results. For example, if data is collected across the exponential and stationary phase, the doubling-time will appear to be much longer than the true value.

lineage specification. Alternatively, the increased proliferation rate may result from the change in medium. When placed in withdrawal conditions, ES cells stop receiving the extrinsic signals that keep them in the naïve state. It is possible that the cells are settling into a new environment where their doubling time is shorter. Especially when one considers the withdrawal of the Mek/Erk inhibitor PD03, which disrupts Fgf4 signaling; a vital pathway in cell growth (Lanner and Rossant, 2010; ter Huurne et al., 2017). If either of these is the case, then why is it that no difference in doubling time was observed for 2i ES cells and N2B27 cells (both $\approx 12 hours$)? Chapter 5 uses long-term single-cell imaging to explore this question.

Chapter 4

Data-Embedded Lineage-Tree Analysis

4.1 Introduction

Long-term single-cell imaging and tracking provides access to the genealogical information which is inaccessible to population averaging and single-cell snapshot data. This highly interdisciplinary technique requires several methodologies to be in place to efficiently assess this information. Platforms for imaging live cells for days at a time are becoming more accessible, however there is still no standard protocol for performing and quantifying long-term single-cell tracking. In this chapter I discuss my approach to single-cell imaging, tracking, and quantification. This includes the cell lines that were constructed, the tracking software that was developed and employed, and the library of functions that has been constructed to analyze these data.

4.2 Construction and Validation of a Traceable Cell Line

The efficiency of long-term single-cell tracking can be improved by insertion of a tracking reporter that fluorescently labels a sub-cellular structure. This is different from a reporter of interest which is being quantified to examine a cellular process. The purpose of the tracking reporter is strictly for visualization and segmentation of individual cells.

The tracking reporter must satisfy certain criteria. First, the excitation and emission (ex/em) ranges should be significantly different than that of the fluorophore of interest. This prevents excitation of the tracking reporter from interfering with measurements of the fluophore of interest. Here, the fluorophore of interest is Rex1-GFPd2 , so the red fluorescent protein mCherry was selected for the tracking reporter (Fig. 4.1).

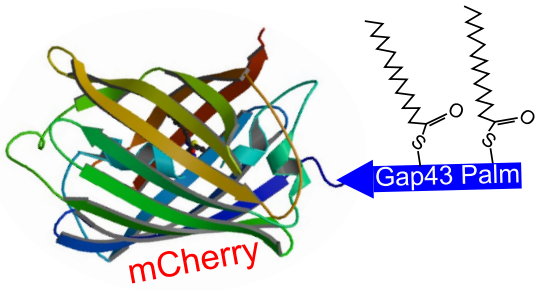


Fig. 4.1 Ribbon Diagram of Gap43-mCherry Protein Structure. mCherry is N-terminally tagged with the palmitoylation sequence of Gap43. Cysteines 3 and 4 are the residues of Gap43 which are modified with palmitate which produces an mCherry that will be attached to the interior of the plasma membrane (Goedhart et al., 2012). The structure of mCherry was modified from the RCSB Protein Data Bank (Berman et al., 2000).

Second, the tracking reporter should create a bounding surface for the fluorophore of interest, effectively bundling cell segmentation and fluorphore quantification. Rex1-GFPd2 is located in the cytosol, meaning that the bounding surface is the plasma membrane. Goedhart et al., 2012, demonstrated in HeLa cells that N-terminal tagging of a flurophore with the Gap43 palmitoylation sequence is sufficient to label the plasma membrane, which is what I have done with mCherry.

The sequence for the Gap43-mCherry fusion protein was transfected by lipofection into Rex1-GFPd2 ES cells and was randomly inserted by PiggyBac integration. Subclones were then selected for their brightness and propagated (Fig. 4.2). The Gap43-mCherry reporter does indeed label the plasma membrane. This is especially apparent when cells come into contact with one another. Although some cytoplasmic signal of Gap43-mCherry can be seen, in what is perhaps the endoplasmic reticulum or some other transport vesicles, this labeling is satisfactory for the purposes of manual cell segmentation.

The kinetics of the new Rex1-GFPd2/Gap43-mCherry ES cell line was validated against the parent Rex1-GFPd2 ES cell line on the population-level (Fig. 4.3). A subclone that showed no significant difference from the parent line was selected, propagated, and used for all subsequent experiments (ANOVA with Holm-Sidak post-test, p-value = 0.6892)).

4.2.1 Summary

The Gap43-mCherry tracking reporter was constructed and transfected into the Rex1-GFPd2 ES cell line. It was shown here that ES cells are capable of expressing the Gap43-mCherry fusion protein, which locates to and labels the plasma membrane. The Rex1-GFPd2/Gap43-

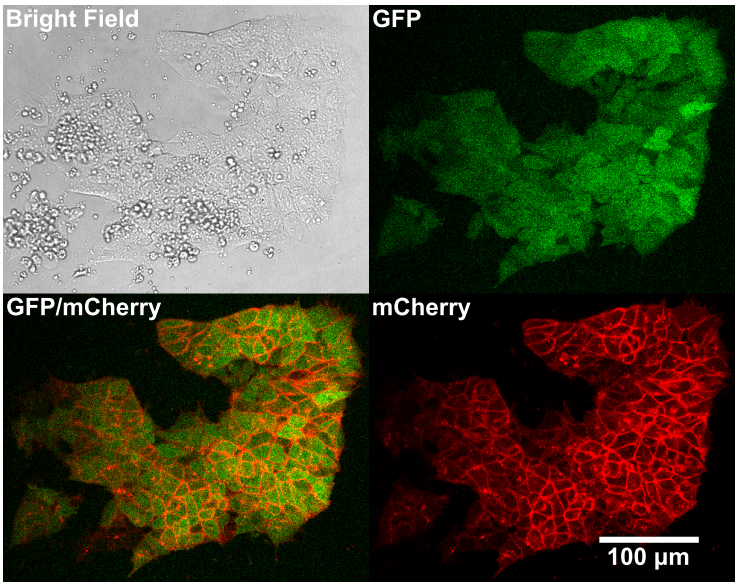


Fig. 4.2 The Rex1-GFPd2/Gap43-mCherry Dual Reporter ES Cell line. The addition of a tracking reporter, such as Gap43-mCherry, provides increase the traceability of cells in dense clusters. Images of the dual reporter Rex1-GFPd2/Gap43-mCherry ES cells in 2i on laminin. Shown is the bright field in gray (top left), the Rex1-GFPd2 in green (top right), the Gap43-mCherry in red (bottom right), and a merge of the Rex1-GFPd2/Gap43-mCherry (bottom left). The Rex1-GFPd2 ES cell line was modified by randomly inserting the Gap43-mCherry sequence to provide better demarcation of the cell boundaries.

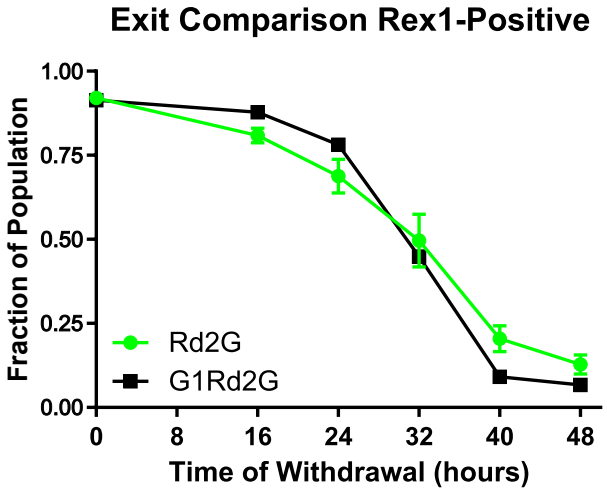


Fig. 4.3 Exit Kinetics Validation for Rex1-GFPd2/Gap43-mCherry ES Cells. Comparison of exit kinetics upon withdrawal of 2i between the parental Rex1-GFPd2 ES cell line (Rd2G, green line) and its subclone Rex1-GFPd2/Gap43-mCherry ES cell line (G1Rd2G, black line). The exit kinetics of both cells lines were determined by conducting the the population-level assays detailed in Sec. 3.2.1. ANOVA with Holm-Sidak post-test shows no significant difference in kinetics (p-value = 0.6892, n = 2).

mCherry ES cell subclone was selected on the basis of brightness and validated by population-level exit kinetics.

4.3 Long-Term Single-Cell Imaging and Tracking

Long-term single-cell imaging was conducted by confocal microscopy for three conditions. The transition to exit from naïve pluripotency was observed using Rex1-GFPd2/Gap43-mCherry ES cells in N2B27 and two control conditions were used to establish distributions for the positive and negative status of Rex1-GFPd2 expression levels. The positive control was Rex1-GFPd2/Gap43-mCherry ES cells and the negative control was Gap43-mCherry ES cells without the Rex1 reporter, both cultured in 2i. The positive control was also used as a baseline for ES cell behavior.

Single cells were plated on laminin in 2i at single cell density (Fig. 4.4). After 12 hours, the medium was aspirated and replaced with N2B27 for exit conditions or 2i for control conditions. Multiple positions were selected for imaging and confocal microscopy was used to generate z-stacks, with non-overlapping optical sections, for GFP, mCherry, and bright field.

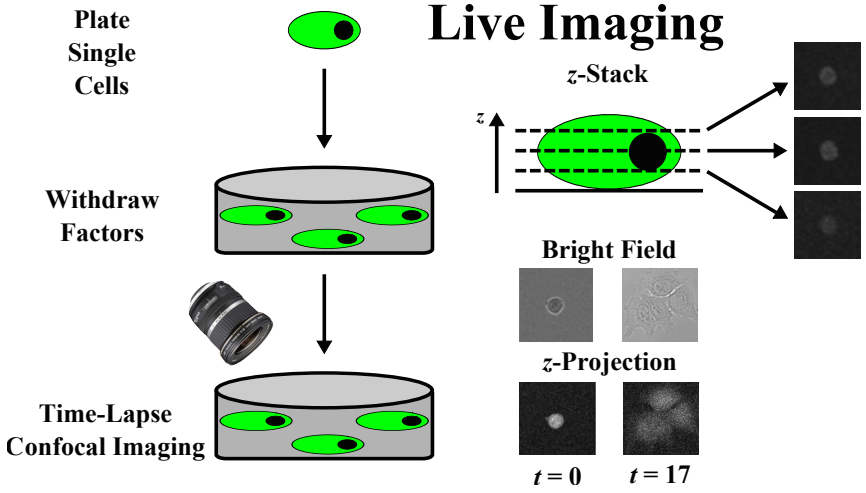


Fig. 4.4 Long-Term Single-Cell Imaging Assay. ES cells were plated on laminin at single-cell density (500 cells/10 cm^2) in 2i. Rex1-GFPd2/Gap43-mCherry were used for the positive control and exit conditions and E14 IVC ES cells with Gap43-mCherry were used for the negative control. After 12 hours, the medium was aspirated and replaced with N2B27 for the exit conditions or 2i for the positive and negative control. The cells were allowed to recover in 5% O_2 incubator for one hour before time-lapse confocal microscopy of multiple positions. Z-stacks were collected for bright field, GFPd2, and mCherry. Cells were imaged every 30 minutes. Further details can be found in Sec. 2.3.

The set of images from each position was manually inspected. The slice that provided the best qualitative images were extracted from the mCherry and bright field channels. The slices of the GFP channel were summed together to amplify the signal of the single allele Rex1-GFPd2. Attempts using different wide field microscopes produced either poor quality images for tracing or low signal-to-noise-ratios for GFPd2. Channels were then concatenated and were exported from imageJ with the Bio-Format plugin as 'OME-TIFF' files.

Three methods were explored for single-cell quantification and lineage reconstruction. Fully manual tracking was cumbersome and quickly proved to be unscalable. Two semi-automated tracking tools were investigated. An earlier version of The Tracking Tool (TTT) was unsatisfactory (Hilsenbeck et al., 2016). This version, which predates the current release (unavailable at that time), had rudimentary functionality, manual segmentation was restricted to regular ellipses, and there was no means to export data. The software that was used for this thesis comes from Blanchard et al., 2009.

Tracer is a semi-automated cell tracking tool that was used for the analysis of the image hyperstacks (xytc). The manual aspect of this software is cell identification and segmentation. Cells were identified by visual inspection of the different channels. Segmentation was performed free-hand by outlining the image of the tracking reporter with the mouse cursor. Once the outline had been complete a digital mask was automatically generated over the outlined cell image, indicating that it had been tracked for that frame. Cells were also prescribed one of three fates: 'divided', 'died', or 'unknown'. The 'unknown' cell fate resulted when the cell moved out of frame or could not be identified faithfully. This data was saved and was read by Otracks, the analysis component of the software. This bespoke tracking and analysis software was modified in collaboration specifically for the purpose of this work (software written by G. Blanchard). The Otracks function "stan" automatically tabulates the average fluorescence intensity, cell area, and cell centroid along with genealogical data stored in the cell name (explained in Sec. 4.4). Data from each video was then exported to a text file. These text files were compiled and analyzed using a library of MATLAB functions written by me.

4.3.1 Data Visualization

The tracking data was compiled into individual lineages using a custom built library of MATLAB functions. Construction of the individual lineages are explained in Sec. 4.4. The raw kinetics for each lineage were visualized as a time-series and then manually inspected for fidelity to the raw images.

The following series of figures come in sets of two. The first is a montage of images which follows cells through a branch-trace of the lineage tree (a single line from root to tip), indicated by colored arrows. These show images of bright field, Rex1-GFPd2, Gap43-mCherry, and a merge of GFPd2/mCherry. The second is the set of time-series figures which show graphs of the quantified primary data (position, size, signal) and how they are used to generate secondary data. Colored curves correspond to indicated cells in the montage and the vertical black lines correspond to the time-points of the montage. Although these are only representatives, they highlight general trends that will be rigorously examined in Ch. 5.

Figure 4.5 shows the montage for a representative of the positive control, ES cells in 2i. Cells in the colony appear to remain relatively compact and the expression levels of Rex1 appear to remain relatively constant throughout the duration of the time-lapse. Notice how the Gap43-mCherry tracking reporter shows defined separation between individual cells. Another notable feature is seen in the indicated cell of the 7.5 hour time-point. It is shown "rounded-up", this occurs just prior to cell division.

The graphs of Fig. 4.6 show the quantification of those images. At the end of each cell-cycle, the mean fluorescence can be seen to spike. This corresponds with the pre-division "rounding" event. Also, it appears that, over subsequent generations, mean Rex1-GFPd2 intensity increases as cell area decreases. The high variability in GFP levels is mitigated by taking the product of these two values to calculate total fluorescence. This produces curves with seasonal Rex1-GFPd2 expression levels over each cell cycle; going from low to high. Note that total cellular Rex1-GFPd2 shows a seasonality in expression level which corresponds to the cell cycle. This is expected, because as the cell becomes larger, the absolute number of molecules must increase to maintain the same concentration (Filipczyk et al., 2015; Padovan-Merhar et al., 2015). Last, when the average velocity is calculated from one time-point to the next, cells appear to be most active when they are first born and at division.

Figure 4.7 shows the montage for a representative of ES cells differentiating in N2B27. The cells appear to be less tightly joined than the positive control and cells can be seen to emigrate from the colony starting at 26.0 hours. In that same frame, simultaneous pre-division "rounding" can be seen in sister cells, indicated by arrows. The Rex1-GFPd2 signal appears to be "on" in all cells at 26.0 hours, and then they all seem to shut it "off". This is more easily seen in the total Rex1-GFPd2 signal in Fig. 4.8. Both divisions and Rex1-GFPd2 down-regulation show a high level of synchronicity. Also, ES cells transitioning to exit appear to be larger and and more motile than ES cells in 2i.

Figures 4.9 and 4.10 show the data from the negative control, ES cells in 2i which do not express the Rex1-GFPd2 reporter. As expected, the negative control recapitulates the positive control, as it should, with the exception that there is no Rex1-GFPd2 expression. All that can be seen in the GFP channel is background noise. The resulting trajectories of Rex1-GFPd2 total fluorescence signal for all tracked cells for each condition and replicate are shown in Fig. 4.11. These positive and negative controls were used to perform the thresholding described in Sec. 2.3.4.

Finally, Table 4.1 summarizes the scale of the single-cell tracking. This experiment has been performed in two independent biological replicates, which differ in number of data-points. I was often able to trace lineages to the fourth and fifth generations. The difference in the number of cells traced comes from a difference in the size of the field of view. Replicate 2 was conducted first and many cells were lost out of view. As a result, the field of view was doubled for Replicate 1 which resulted in more traceable cells. Although the field of view was doubled, the resolution and exposure time of imaging was maintained. The negative control, E14 IVC in 2i, show that 6 clones were traced for both Replicate 1 and 2. However, Replicate 2 has twice as many data-points. This is because many of the colonies in Replicate 1 rounded up into domed colonies and individual cells became untraceable.

Table 4.1 Summary of long-term single-cell data. Shown are the number of cells, the number of clones, and the number of data-points tracked for all conditions in both biological replicates. 'Cells' indicate the number of individual cells tracked for any period of time, 'Clones' indicate the number of lineages tracked from a single cell, and 'Data-Points' indicate the total number of recordings made for each cells.

		Replicate 1			**Replicate 2**		
Cell line	**Condition**	**Cells**	**Clones**	**Data-Points**	**Cells**	**Clones**	**Data-Points**
Rex1-GFPd2	N2B27	1384	64	19,138	573	43	8,691
Rex1-GFPd2	2i	405	17	6,141	154	18	2,438
E14 IVC	2i	56	6	664	52	6	1,258
Total	-	1845	87	25,943	779	67	12,387

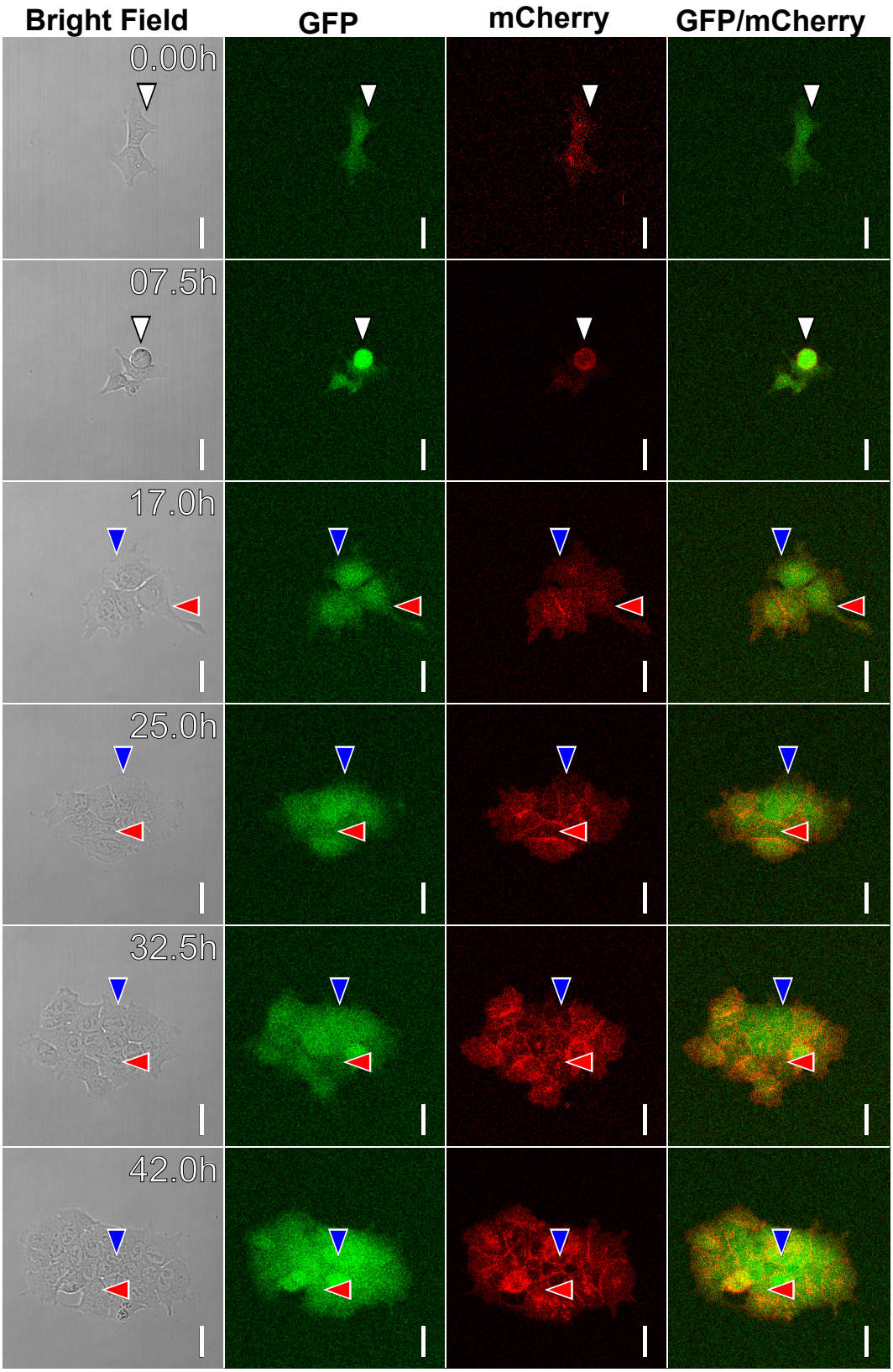


Fig. 4.5 **Positive Control.** Long-Term Single-Cell Imaging Montage of Positive Control. Rex1-GFPd2/Gap43-mCherry ES cells in 2i were used as the positive control for long-term single-cell imaging. They served as a baseline for all inferences about exiting ES cells. This includes fluorescence level, motility, area, and cell-cycle duration. The colored arrows indicate cells belonging to one branch-trace of the lineage tree and correspond to the colored curves in Fig. 4.6. Scale bar $= 25\mu$m.

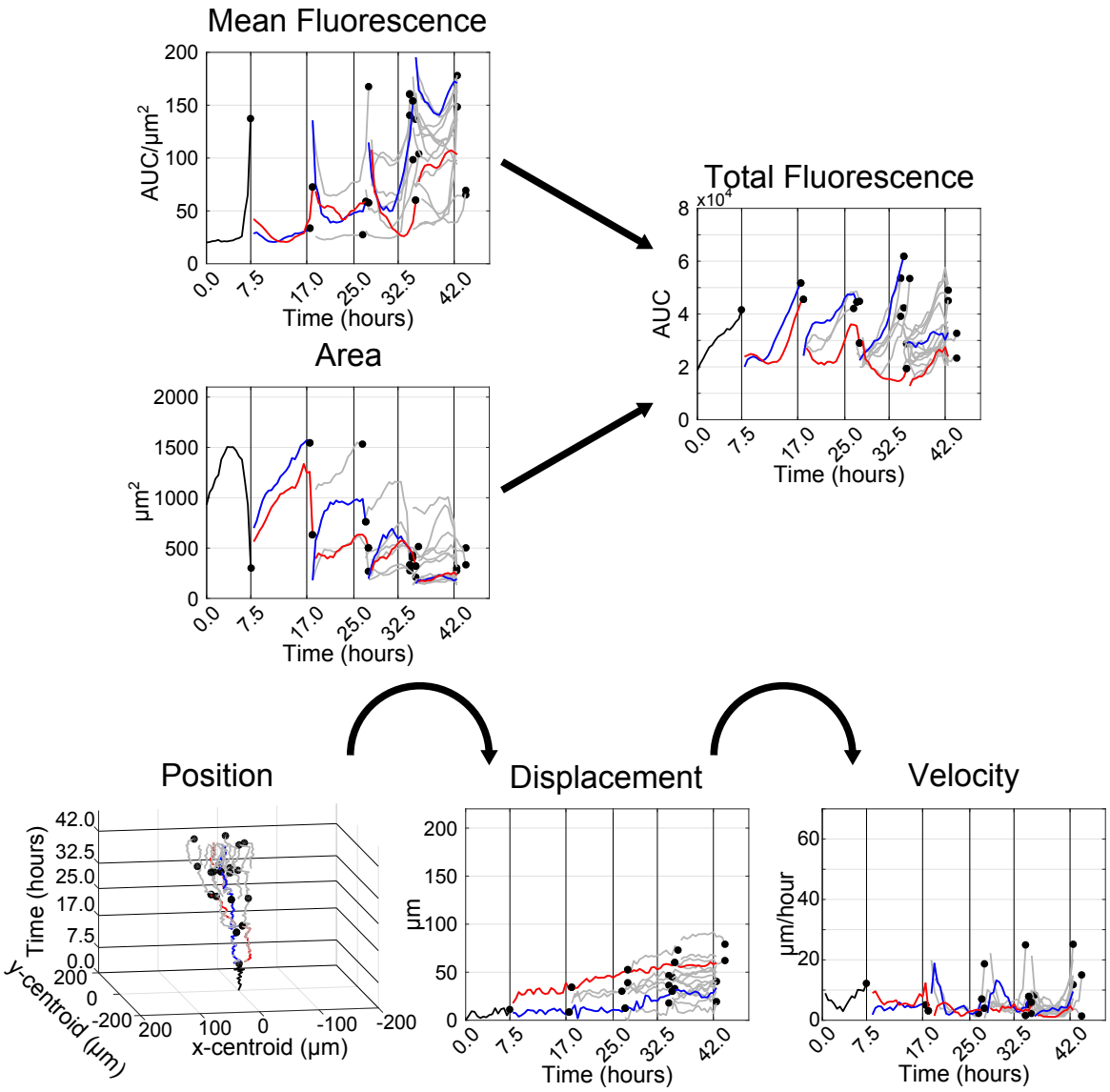


Fig. 4.6 Quantification of Long-Term Single-Cell Imaging for Positive Control. Representations of primary data from the full time-course of one clonal lineage and how they are used to generate secondary information. The vertical black lines correspond to the time points in the montage from Fig. 4.5 and the colored curves correspond to the cells marked with arrows.

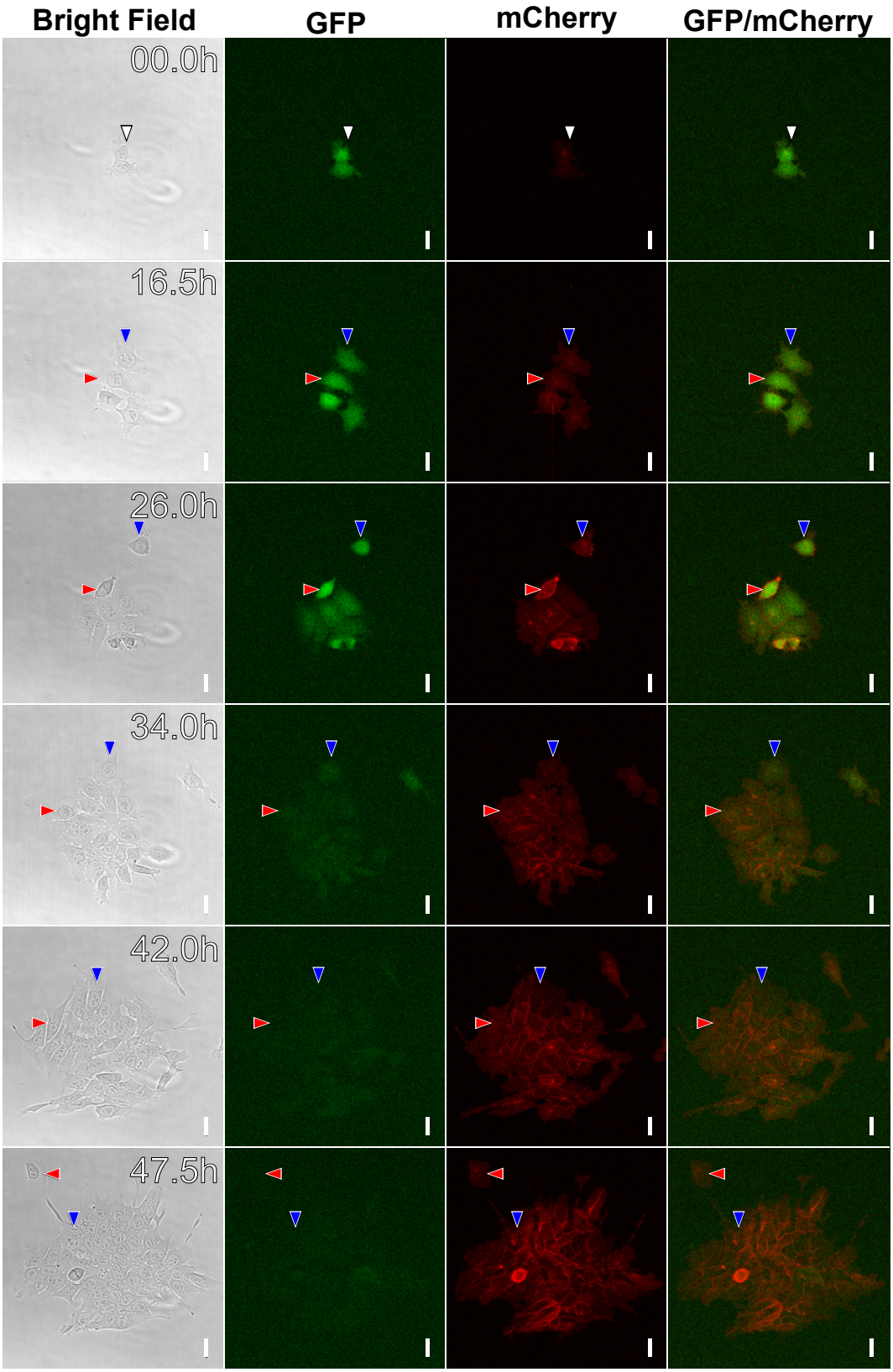


Fig. 4.7 **Exit Conditions.** Long-Term Single-Cell Imaging Montage of Exiting ES Cells. Rex1-GFPd2/Gap43-mCherry ES cells differentiating in N2B27 are compared against the positive control to make inferences about how ES cells change during the differentiation process. The colored arrows indicate cells belonging to one branch-trace of the lineage tree and correspond to the colored curves in Fig. 4.8. Scale bar $= 25\mu$m. (n.b. the field of view is larger than the other two montages).

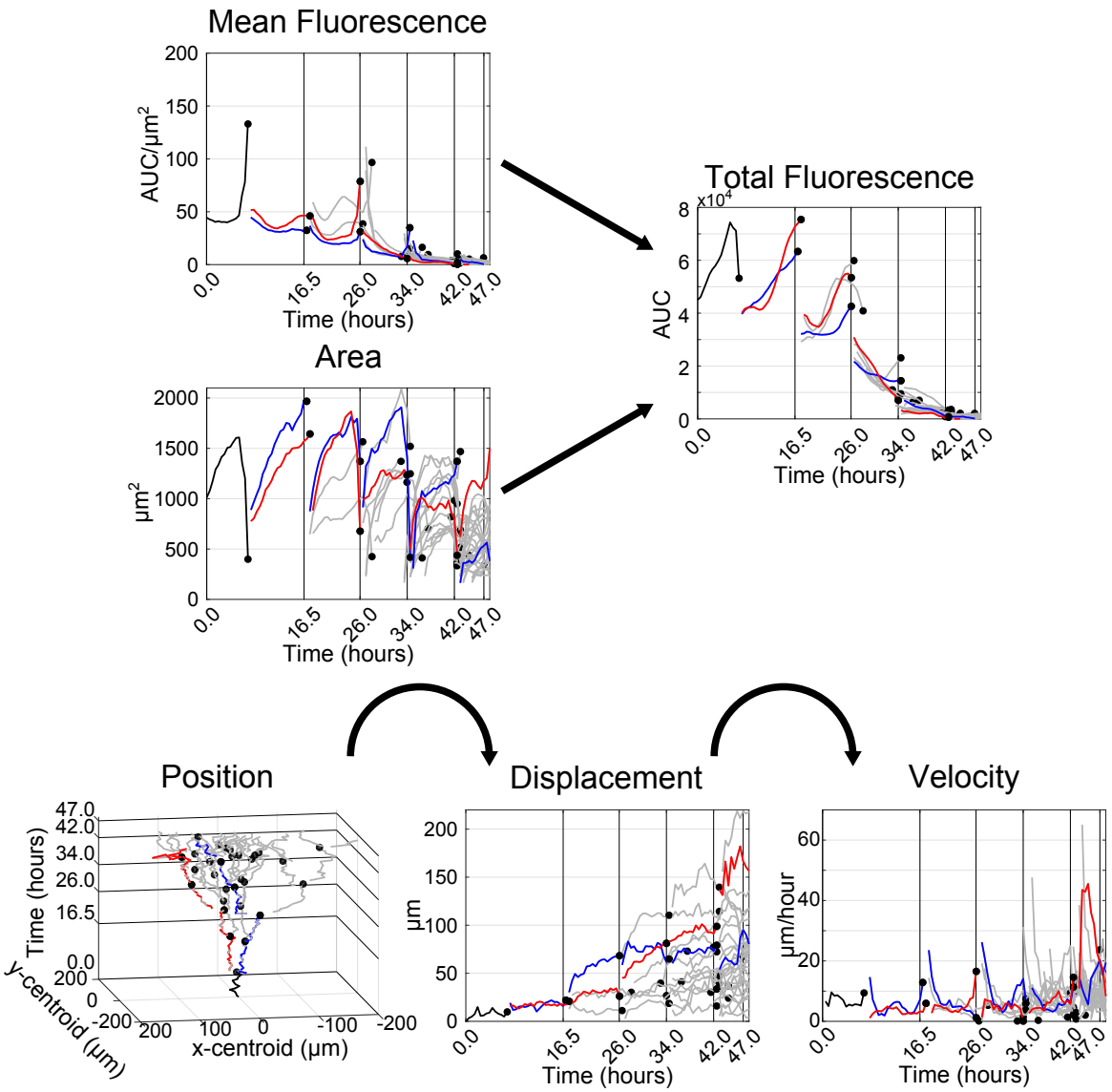


Fig. 4.8 Quantification of Long-Term Single-Cell Imaging for Exiting ES Cells. Representations of primary data from the full time-course of one clonal lineage and how they are used to generate secondary information. The vertical black lines correspond to the time points in the montage from Fig. 4.7 and the colored curves correspond to the cells marked with arrows.

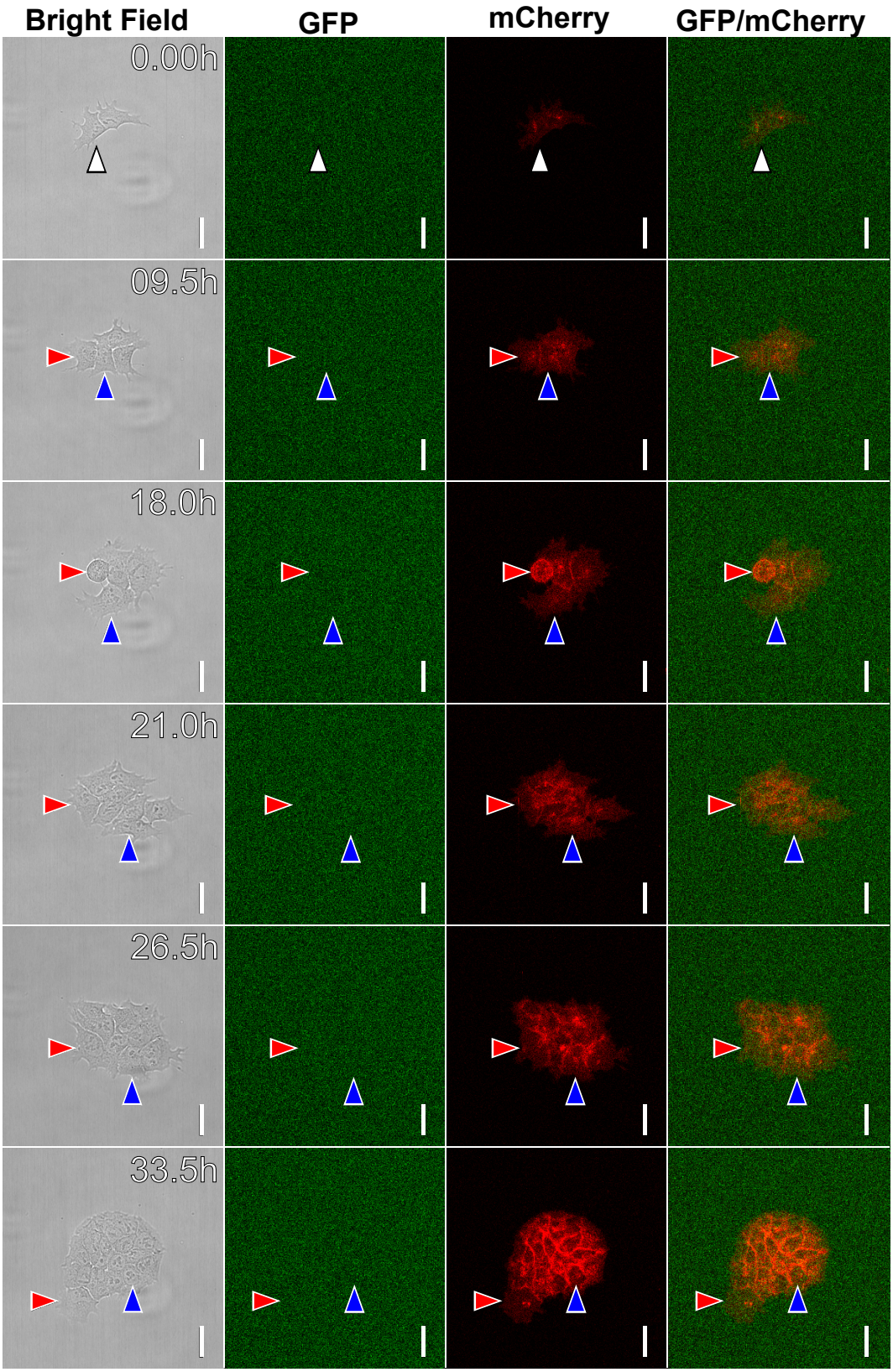


Fig. 4.9 **Negative Control.** Long-Term Single-Cell Imaging Montage of Negative Control. E14-IVC/Gap43-mCherry ES cells in 2i were used to generate the distribution of Rex1-GFPd2 negative cells. The colored arrows indicate cells belonging to one branch-trace of the lineage tree and correspond to the colored curves in Fig. 4.10. Scale bar $= 25\mu$m.

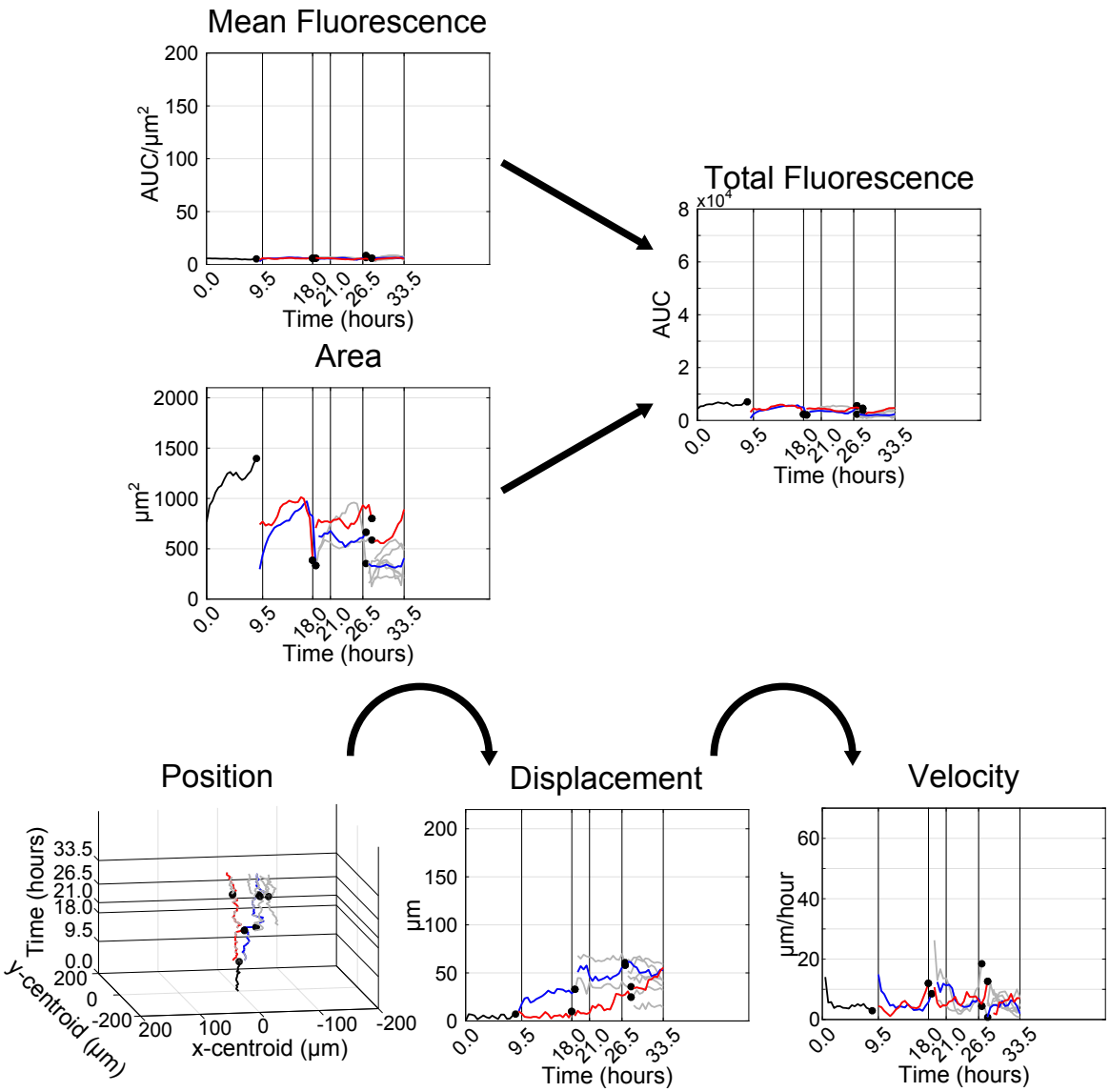


Fig. 4.10 Quantification of Long-Term Single-Cell Imaging for Negative Control. Representations of primary data from the full time-course of one clonal lineage and how they are used to generate secondary information. The vertical black lines correspond to the time points in the montage from Fig. 4.9 and the colored curves correspond to the cells marked with arrows.

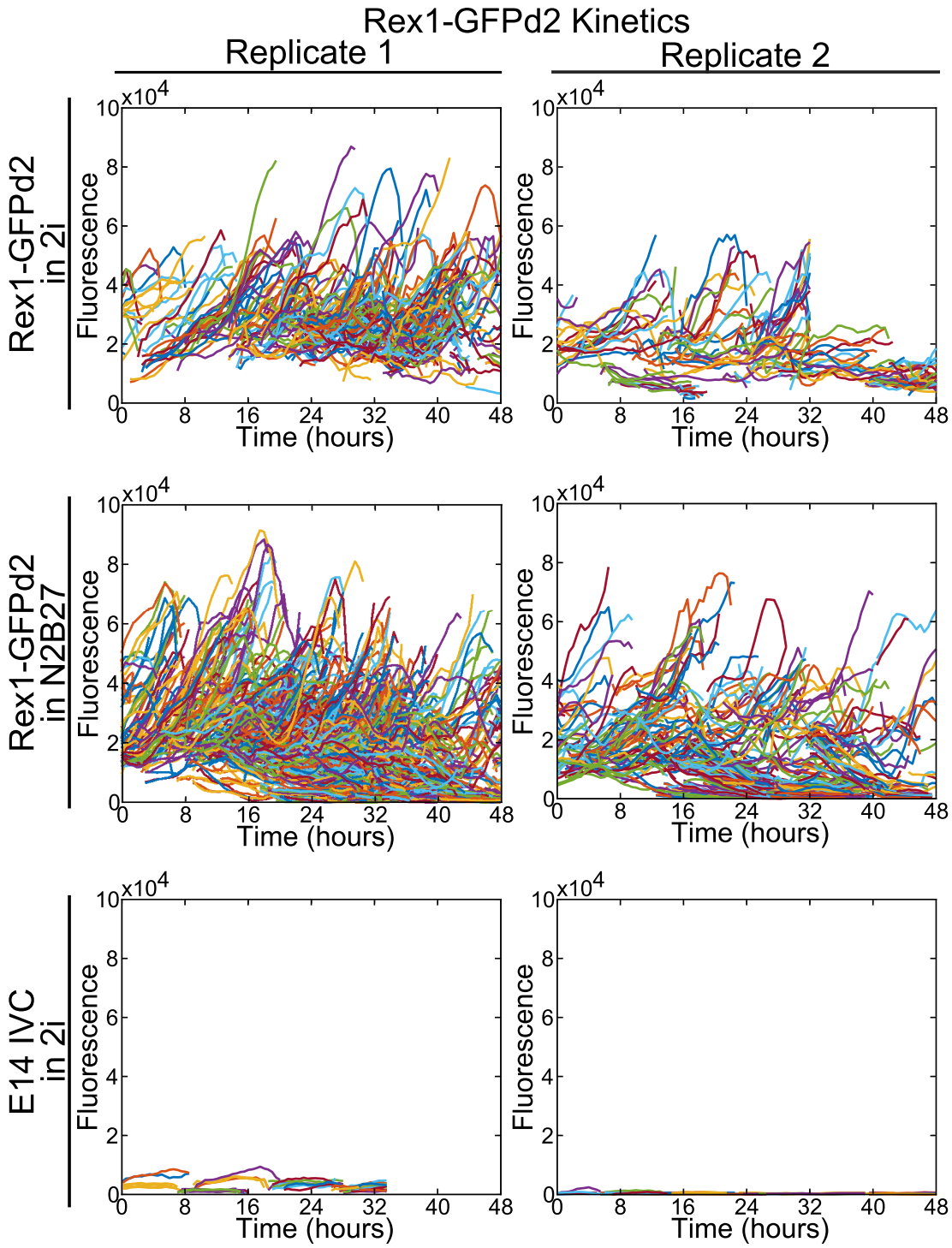


Fig. 4.11 The global real time kinetics of total cellular fluorescence. All cells that have been tracked for the positive control (Rex1-GFPd2 in 2i), exit conditions (Rex1-GFPd2 in N2B27), and the negative control (E14 IVC in 2i) are shown for both biological replicates. Each curve represents an individual cell, line color is arbitrary, and a 5-point rolling average has been applied.

4.3.2 Summary

Time-lapse confocal microscopy was used to track ES cells as they exit naïve pluripotency. Rex1-GFPd2 ES cells and WT E14 IVC ES cells in 2i were employed as the positive and negative controls for the Rex1-GFPd2 expression levels. Existing semi-automated tracking software was used to track cells and quantify the mean fluorescence signal, position, and size. These data were then used to generate secondary data to be used for statistical analysis.

4.4 Data-Embedded Lineage-Tree Structure

The data-embedded lineage-tree is the base structure resulting from single-cell tracking. Lineage reconstruction produces these rooted, bifurcating dendrograms embedded with information at each node. This structure is composed of two main features: genealogy and cellular kinetics. The lineage-tree contains the genealogical information which details the relationships to other cells (mother, sister, cousins), generation number, and the terminal outcome of cells (death, division, or unknown). Each node represents a cell which contains primary cellular information including lifespan, position, molecular densities, neighbors, and size. Secondary data can then be derived from these primary kinetics data. This section outlines the construction and encoding of the data-embedded lineage-tree.

4.4.1 Encoding Primary Data

The first aspect of the data-embedded lineage-tree is the genealogy (Fig. 4.12). Reconstruction of the genealogy for a cellular lineage requires knowing from where a cell originated and its ultimate fate. Blanchard et al., 2009, have implemented a binary string naming system which houses this genealogy. To illustrate with an example, the origin of the founding mother cell for a clonal lineage is unknown, so it is prescribed an arbitrary name. All clonal founder cells are named ‘c’ followed by an integer, which corresponds to the number of cells in the first frame of imaging. In this case it is cell ‘c1’. When ‘c1’ divides, the two daughter cells are named by appending the mother name with either an ‘a’ or ‘b’. So the daughters of ‘c1’ are ‘c1a’ and ‘c1b’. This allows for simple identification of genealogical relationships and the generation to which a cell belongs (table in Fig. 4.12).

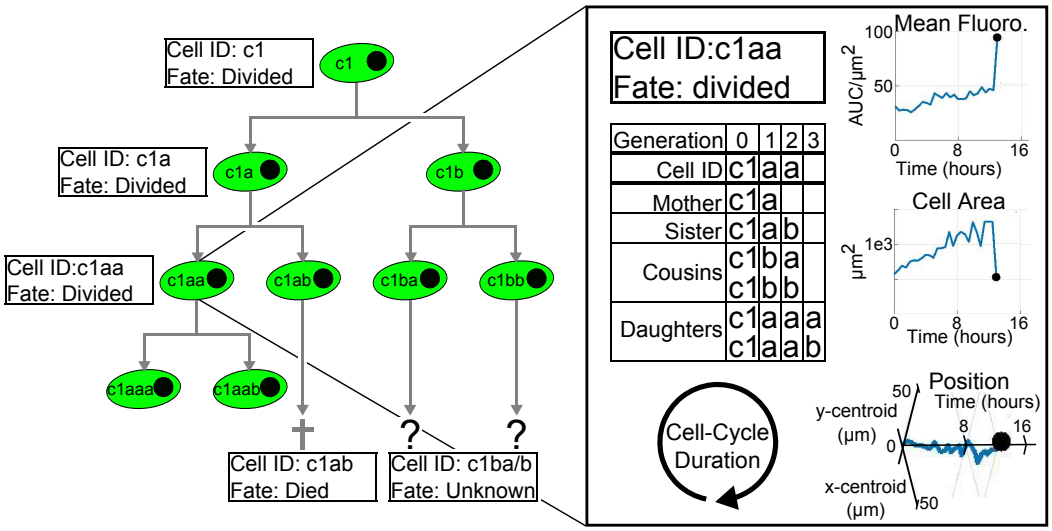


Fig. 4.12 Data-Embedded Lineage-Tree Structure. Lineage reconstruction is performed using the genealogical information of the cell that is stored in its name (CellID) and fate. The cell name contains the generation and all prior genealogical information including ancestors, sister, and cousins. The fate of each cell is 'died', 'divided', or 'unknown'. The 'unknown' fate occurs when the cell is either lost out of view or indistinguishable in the colony. Each cell houses the primary kinetics data for mean cellular fluorescence, cell area, cell position and cell-cycle duration.

Cells must also be prescribed a fate. The binary 'ab' nomenclature is effective in generating the structure of the lineage tree, but it does not detail the fate of cells at branch terminals. The cell could have died, migrated out of the frame of view, or it could have still been in frame at the termination of imaging. All cells have been prescribed one of three fates: 'division', 'death', or 'unknown'. The case of the 'unknown' fate houses two cases, migration and termination of observation.

The second aspect is the cellular attributes data. Each node is embedded with the kinetics from all observable attributes. These have been stored in individual one dimensional vectors of equal length composed of floating point numbers. Recorded attributes include the times that a cell was alive ('t'), the x- and y-centroid ('xc' and 'yc'), cell area ('area') and the mean cellular fluorescence signal ('f').

Table 4.2 shows how the features of the data-embedded lineage-tree were encoded. The essential genealogical attributes are the cell name ('id') and the cell fate. From these two pieces of information, all other genealogical relationship can be determined. I have chosen to make cell fate a string for ease of human understanding, but this could have alternatively been the set {0,1,2} for the three different fates.

Table 4.2 Encoding the Data-Embedded Lineage-Tree. The 'attributes' for each 'cell' of a 'lineage' are individually encoded under the cell field name. Associated with each attribute is the field name, definition, and data type. Encoding 'optional' secondary attributes are not essential, as they can be determined *in situ*, however this can make certain analyzes easier for the user.

Structure Fields					
Lineage		**Cell**	**Attribute**	**Definition**	**Type**
lineageName	.	cellName	. id	cell name	string
			. fate	cell fate	string
			. t	time	float array
			. xc	x-centroid	float array
			. yc	y-centroid	float array
			. area	segmentation area	float array
			. f	mean fluorescence	float array
(Optional)			. parent	parent name	string
			. sister	sister name	string
			. generation	cell generation	integer
			. tf	total fluorescence	float array

Secondary optional attributes to encode, which are helpful for analysis, are the names of mother ('parent') and sister ('sister') and the generation number ('generation'). It is also beneficial to encode secondary attributes which are central to the analyzes being performed. In this study, I have included the total cellular fluorescence of Rex1-GFPd2 signal.

4.4.2 Calculations of Secondary Data

The secondary data is derived *in situ* from the primary data. Cell-cycle duration ('life') was calculated by taking the difference of the first and last time that the cell was observed. The total cellular fluorescence was discretized in several ways. The initial ('tfi') and final ('tff') fluorescence was determined by taking the mean of the first and last three time-points of the total fluorescence, similar to Filipczyk et al., 2015. The average rate of change ('dtf') in fluorescence over the cell's lifetime was determined by taking the difference between the final and initial fluorescence and dividing that by the cell-cycle duration. The lifetime fluorescence accumulation ('tfaccum') was calculated by numerical integration of the total fluorescence using the trapezoidal method. Table 4.3 shows the calculations and restrictions for all remaining secondary attributes, including those pertaining to fluorescence signal.

It is important to consider which attributes can be used for which cells. For instance, it does not make sense to calculate the cell-cycle duration for cells that were not seen both

Table 4.3 The secondary attributes are shown with their field name and how they were calculated. The 'Condition' specifies which events must be observed for that attribute to be valid.

Secondary Attribute	**Field Name**	**Calculation**	**Condition**
cell-cycle duration	life	$\mathrm{t}(end)-\mathrm{t}(1)$	born & divided
initial fluorescence	tfi	mean[$\mathrm{tf}(1:3)$]	born
final fluorescence	tff	mean[$\mathrm{tf}(end-2:end)$]	divided
mean lifetime fluorescence	tfbar	mean[tf]	born & divided
rate of fluorescence accumulation	dtf	$\frac{\mathrm{tff}-\mathrm{tfi}}{\mathrm{life}}$	born & divided
lifetime accumulated fluorescence	tfaccum	trapz[t,tf]	born & divided
initial area	ai	mean[$\mathrm{area}(1:3)$]	born
final area	amax	max[area]	born & divided
mean area	abar	mean[area]	born & divided
displacement	r	$\sqrt{(\mathrm{xc}-\mathrm{xc}_i)^2+(\mathrm{yc}-\mathrm{yc}_i)^2)}$	none
average speed	vbar	$\frac{\sum\sqrt{(\mathrm{x}_i-\mathrm{x}_{i-1})^2+(\mathrm{y}_i-\mathrm{y}_{i-1})^2)}}{\mathrm{life}}$	born & divided

to be born and divide, e.g. the founding cell of a clone was not seen born or a cell that migrated off screen was not seen to divide. This is what the conditions in Table 4.3 are used to determine. If an attribute requires information from the initial stage of the cell cycle, like the initial fluorescence, then the cell birth must be witnessed. Similarly, if an attribute requires information from the end stage of the cell-cycle, such as the final fluorescence, then the cell division must be witnessed. Or if the attribute requires information from the lifetime of the cell, such is the case with cell-cycle duration, then both events must be witnessed. Otherwise, the missing information will skew the results.

Alg. 1 shows how attributes were indexed for analysis. All attributes for cells were calculated and placed into attribute vectors with corresponding vectors for generation number and the cell fate (divided, died, unknown). The generation and fate vectors were used to index the cells with observed births and divisions. Attributes were sorted into subsets for the different observation conditions and the index vectors were used to retrieve the cells which fit that criteria. These data can also be paired with generation number to make generational distributions of different attributes. Discretizing time into generations is one method that has been used in this thesis to examine the time evolution of cellular behaviors.

Algorithm 1 Cell-exclusion for attributes

Given a set of attributes $A = \{a_1, a_2, \ldots, a_n\}$,
s.t. vectors $a_p \in \mathbb{R}^m$ house the values of the attribute for m cells.
generation = lineageName.cellNames.generation ▷ vector of generation numbers
fate = lineageName.cellNames.fate ▷ vector of cell fates

Index cells
bornIndex = (generation > 0) ▷ cells with observed birth
divideIndex = (fate == 'divided') ▷ cells with observed division
Where bornIndex & divideIndex are logical vectors

Sort attributes into subsets
$a^{(0)} = \{a_1^{(0)}, \ldots, a_i^{(0)}\}$ ▷ observed birth required
$a^{(1)} = \{a_1^{(1)}, \ldots, a_j^{(1)}\}$ ▷ observed division required
$a^{(2)} = \{a_1^{(2)}, \ldots, a_k^{(2)}\}$ ▷ observed birth & division required
$i, j, k \in \mathbb{N}$, s.t. $i + j + k = n$

Retrieve data
$A_0 = \{a^{(0)}(\text{bornIndex})\}$
$A_1 = \{a^{(1)}(\text{divideInedx})\}$
$A_2 = \{a^{(2)}(\text{bornInedx + divideIndex})\}$

4.4.3 Summary

Standardized raw data and statistical analyzes have yet to be established in long-term single-cell imaging. Therefore I have used this section to explain how I have managed these complex data by the construction and encoding of the data-embedded lineage-tree. Extraction of the primary kinetics data allows for the generation of secondary data which can then be visualized for manual inspection and statistical analysis. The raw time-lapse images and single lineage quantification have hinted at interesting behaviors that will be explored in Ch. 5.

4.5 Analytical Methods for the Interrogation of Data-Embedded Lineage-Trees

Once the reconstruction of cell lineages has been encoded into data-embedded lineage-trees, analyzes of a cellular process can be performed using many lineages. There is no standard frame-work for analyzing these forms of data, so each analysis is conducted on an *ad hoc* basis, using various statistical methods (Skylaki et al., 2016). This section outlines the methods and algorithms used for the analyzes conducted in Ch. 5.

4.5.1 How Should Cells be Compared?

Consider the expression levels of a gene for two sister cells which divide at different times (Fig. 4.13 A). A point-wise comparison of expression levels can be made between the two cells until the first cell divides (Fig. 4.13 B). But what then should be done with the remaining information of the cell which has yet to divide? This issue was managed by converting the data into discrete attributes (Fig. 4.13 C). This is what has been done for encoding secondary information of data-embedded lineage-tree.

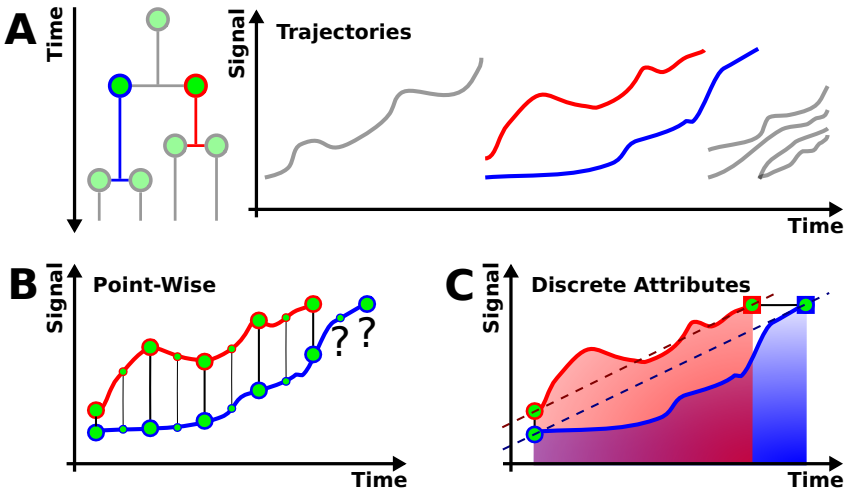


Fig. 4.13 Comparing the Kinetics of Two Cells. Given a lineage (A) with sister cells that divide at different times (red and blue), how should their genetic trajectories be compared? A point-wise comparison is not sufficient (B) because this leaves a data over-hang for one of the cells. One method is to (C) discretize the signal into attributes such as initial (circles) and final signal (squares), rate of production (dotted lines), and total lifetime production (AUCs).

Once these pairwise attributes have been generated how should these data be compared? One important question in genealogical analysis is determining levels of similarity. Pearson's

correlation has been used to compare the cell-cycle duration in different systems (Cannon et al., 2015; Herberg et al., 2016; Sandler et al., 2015; Scherf et al., 2012). The Pearson correlation can determine which genealogical relationships are more highly correlated. But it cannot determine if they are the same, e.g. sets of cells can be correlated and cycle at different rates. A paired t-Test can determine if two distributions share the same mean, but it does not provide a method to compare the levels of similarity between different relationships. I suggest using the paired t-Test to determine if a genealogical relationship is the same and only then employ the Pearson's correlation to compare the levels of similarity across relationships.

The significance for the Pearson correlation was determined using permutation testing (Alg. 2). The correlation for a relationship was calculated using MATLAB's $corr(\cdot)$ function. A distribution for the test statistic ρ was generated by permuting the data and calculating the correlation coefficient for each permutation. This distribution was then used to calculate the p-value by determining how many correlation coefficients from the permutations were more extreme than the observed correlation.

Algorithm 2 Permutation testing for significance

Given $M_0 \in \mathbb{R}^{m\times 2}$ ▷ cols are genealogical relationship
$\rho = corr(M_0(:,1), M_0(:,2))$ ▷ Pearson's correlation coefficient

$M_0 \in \mathbb{R}^{m\times 2} \rightarrow M \in \mathbb{R}^{2m\times 1}$ ▷ stack matrix into vector
$P = zeros(n,1)$ ▷ storage vector for random Pearson's correlation coefficients

for i = 1 : n **do** ▷ commence permutations
    permM = randperm(M) ▷ random permutation of data
    $P(i) = corr(permM(1 : n/2), permM(n/2+1 : end))$ ▷ store random ρ in P
end for

if Upper-Tailed **then** ▷ determine number of extreme values from permutations
    $extreme = sum(P \geq \rho)$
else if Lower-Tailed **then**
    $extreme = sum(P \leq \rho)$
elseTwo-Tailed
    $extreme = sum(P \geq \rho || P \leq \rho)$
end if

p–value = $\frac{extreme}{n+1}$ ▷ sensitivity dictated by number of permutations

Random permutations were carried out when the number of unique permutations exceeded the number required for the desired significance. This was often the case. Otherwise, all unique permutations were used. E.g., If the desired significance was a p-value = 0.00001, then 100,000 permutations need to be correlated. If there were 10 or more pairs of cells, then there were at least 3,628,800 unique permutations. Thus, generating 100,000 random permutations was sufficient. However, if there were only 8 pairs of cells, then there were only 40,320 unique permutations, so double counting would occur with random permutations. Also, this means that the maximal significance is p-value = 1/(40,320) $\approx$ 0.000025.

4.5.2 Genealogical Comparisons

Genealogical comparisons were performed by generating $n \times 2$ vectors where each row held a pair of cells and the columns were the relationship. Unidirectional pairing for intergenerational comparisons were performed in a top down fashion and intrageneration comparisons were made from left to right (4.14). This ensured that all comparisons were made, and that no comparison was made twice (e.g. c1a→c1b then c1b→c1a).

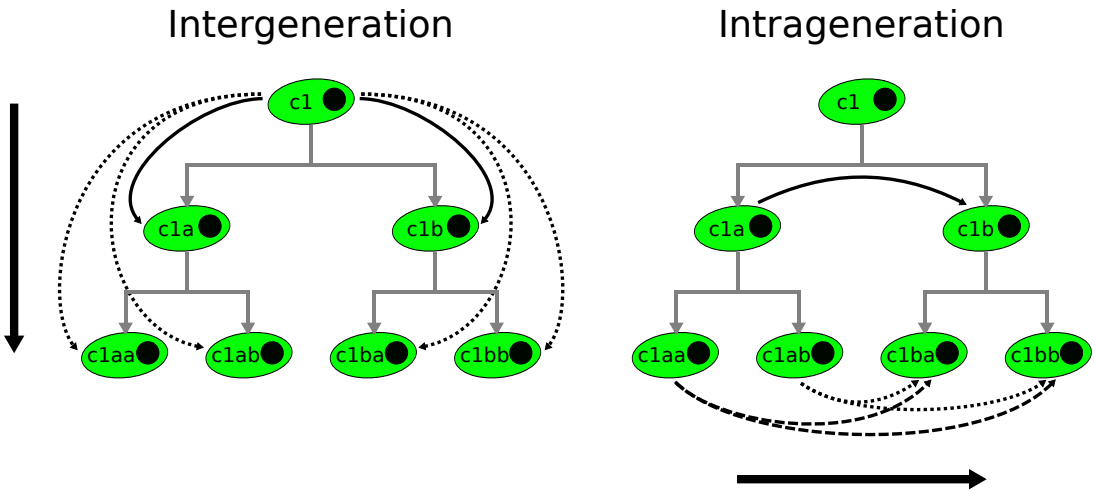


Fig. 4.14 Genealogical Comparisons. Unidirectional genealogical comparisons to prevent double counting. (Left) Intergenerational comparisons are made from the top down; mother to daughters (solid line) and grandmother to granddaughters (dotted lines). (Right) Intrageneration comparisons are made from 'a' cells to 'b' cells. The '…a' sister is compared only against the '…b' sister, and '…a(·)' first-cousins are compared against the the '…b(·)' first-cousins.

Alg. 3 outlines how top-down intergenerational comparisons were conducted. The binary 'ab' nomenclature is used to find all descendant cells D generations away, where daughters are one generation away and granddaughters are two generations away. This is accomplished

by finding all cells that have the same root name as the cell in question and all cells D generations away.

Algorithm 3 Intergenerational Analysis

```
1:  Given comparison distance D ∈ ℕ                          ▷ number of generations away
2:  cellNames = fieldnames(lineageName)                      ▷ names of all cells
3:  generation = lineageName.cellNames.generation            ▷ generation number of all cells
4:  a = lineageName.cellNames.attribute                      ▷ some attribute of all cells
5:
6:  compMat = [];                                            ▷ initiate comparison matrix
7:
8:  n = numel(cellNames)
9:  for i = 1 : n do                                         ▷ step through all cells in lineage
10:
11:     strLength = numel(cellNames(i))                      ▷ get lenghth of binary cell name
12:
13:     nameIndex = (cellNames( :, 1 :strLength) == cellNames(i))   ▷ cells w/ same root
14:     genIndex = (generation == (generation(i) + D))       ▷ cells at desired distance
15:     descendIndex = (nameIndex + genIndex == 2)           ▷ index for relatives of interest
16:
17:     comparisons = [repmat(a(i),sum(descendIndex),1), a(descendIndex)]   ▷ pair cells
18:     compMat = [compMat; comparisons]                     ▷ update comparison matrix ∈ ℝ^{m×2}
19:
20: end for
```

Constructing the comparisons for intragenerational relationships is similar (Alg. 4). The genealogical distance D is the number of generations to the closest common relative; sisters are one generation apart, related by common mother, and first-cousins are two generations apart, related by the grandmother. The first step is to ensure the relationship exists. For instance, after the first division, there are only two generation one sister cells, and no first-cousins. It does not make sense to compare cousins of the first generation because there are none. First-cousins only start appearing in the second generation. Next, unidirectional comparisons are made by considering only cells with an 'a' in the 'generations' place of the string. If sisters are being compared, the terminal character in the string must be 'a'. If cousins are being compared, then the penultimate character in the string needs to be an 'a'. For first-cousins, a cell with the name '1caba' will be passed over, because it has already been compared to '1caaa' and '1caab'. Similar indexing is used to find cells with the same root names and generation number.

Algorithm 4 Intragenerational Comparison

Given comparison distance $D \in \mathbb{N}$ ▷ number of generations away
cellNames = fieldnames(lineageName) ▷ names of all cells
generation = lineageName.cellNames.generation ▷ generation number of all cells
a = lineageName.cellNames.attribute ▷ some attribute of all cells

compMat = []; ▷ initiate comparison matrix

$n = numel$(cellNames)
for $i = 1 : n$ **do** ▷ step through all cells in lineage

  if $D \leq$generation(i) **then** ▷ ensure comparison exists

    if cellNames(i ,$1 : end - D + 1$) == 'a' **then** ▷ only consider 'a'-cells

      rootId = [cellNames(i , $1 : end - D$), 'b'] ▷ root cell name

      nameIndex = (cellNames(:, $1 : end - D + 1$) == rootId) ▷ cells w/ same root
      genIndex = (generation == generation(i)) ▷ cells w/ same generation
      relatIndex = ((nameIndex + genIndex) == 2) ▷ index for relatives of interest

      comparisons = [$repmat$(a(i),sum(relatIndex),1), a(relatIndex)] ▷ pair cells
      compMat = [compMat; comparisons] ▷ update comparison matrix $\in \mathbb{R}^{m \times 2}$

    else cellNames(i ,$end - D + 1$) == 'b' ▷ ignore 'b'-cells

      return 0;

    end if
  end if
end for

Once Pearson's correlation coefficients were calculated and determined to be significant, they could be compared against one another. This was accomplished by Fisher's z-transformation followed by a paired t-Test. Multiple comparison corrections were conducted using the Holm-Sidak method.

4.5.3 Summary

This section has elaborated on the analytical methods used to interrogate data-embedded lineage-trees. I have stepped through the algorithms to access all cell pairs for genealogical comparisons. I have also suggested the use of the paired t-Test in tandem with Pearson's correlation coefficient for the analysis of cell similarity. Finally empirical data was used to estimate the growth curves and exit kinetics. Implementation of the algorithms as MATLAB code and instructions on how to run the code on the data gathered in this thesis will be made available.

4.6 Discussion

Long-term single-cell imaging still holds many challenges (Skylaki et al., 2016). This is somewhat due to the lack of formalism with regard to the technique. In this chapter, I have discussed my approach to imaging, tracking, and analysis. The task of single-cell tracking can be daunting, but some essential molecular and computational tools can enhance the efficiency of the process. A tracking reporter can enhance the raw image analysis. I have labeled the plasma membrane of the Rex1-GFPd2 cell line with the fusion protein Gap43-mCherry. For analysis of the raw images, semi-automated tracking software still appears to be the most reliable tool available. Handling and analysis of this data is still difficult and clunky. I believe that encoding lineage reconstructions into the data-embedded lineage-tree improves the process of analysis. It brings much needed formalism to the often abstract data.

Chapter 5

Single-Cell Analysis of Exit from Naïve Pluripotency

5.1 Introduction

Much of what is known about mouse ES cells, and the exit from the naïve state, comes from snapshots and population data: flow cytometry, RNA seq, and immunostaining (Betschinger, 2017; Betschinger et al., 2013; Kalkan et al., 2017; Kalkan and Smith, 2014; Leeb et al., 2014; Mulas et al., 2017). While this has helped elucidate much about the pluripotency transcription factor network, these techniques are time-destructive and do not preserve genealogical information. Examples of this arise in Ch. 3, e.g. the origin of the asynchronicity in exit could not be determined because the genealogy of cells was not known.

The studies that have employed long-term single-cell tracking to examine ES cells have primarily focused on the serum/Lif system. The mechanism of heterogeneity in pluripotency factors has been attributed to stochastic switching (Glauche et al., 2010a; Herberg et al., 2016, 2014; Herberg and Roeder, 2015), sister-sister genealogical analysis has shown a certain plasticity to the transcription factor network under these conditions (Filipczyk et al., 2015), and elongated cell-cycles have been observed in cells undergoing telomere elongation (Nakai-Futatsugi and Niwa, 2016). So for this work, the technique of long-term single-cell tracking is used to examine the dynamics and the factors affecting the irreversible exit from naïve pluripotency.

5.2 On the Morphology and Motility of ES Cells

Chiron, one of the self-renewal signals in 2i, is a GSK3 inhibitor and as such it stimulates the canonical Wnt signaling pathway by disrupting the β-catenin destruction complex (see Introduction, Fig. 1.7) (Wray et al., 2010; Wu and Pan, 2010). The disruption of the β-catenin destruction complex results in the accumulation of cytosolic β-catenin (Tian et al., 2011). In the context of pluripotency, β-catenin will then translocate to the nucleus where it blocks the repressor Tcf3 (Martello et al., 2012; Wu and Pan, 2010). This allows for the expression of the transcription factor Esrrb, a necessary factor in the core pluripotency circuitry (Martello et al., 2012). Another outcome of β-catenin accumulation is the stabilization of E-caherins and the adherens junctions they form (Tian et al., 2011). This causes cells to adhere to one another, forming compact dome colonies (Nichols and Smith, 2009; Wray et al., 2011). On the other hand, deletion of β-catenin results in diffuse colonies which lack cell-cell contact (Andersson-Rolf et al., 2017; Wray et al., 2011). In this section, long-term single-cell tracking was used to directly observe and quantify cell area and cell speed for N2B27 cells. This was compared against 2i ES cells, which were used as a behavioral base line.

5.2.1 Exit from Naive Pluripotency is Associated with a Slow Decrease in Cell Area

The shape of an ES cell changes substantially through the cell-cycle. Daughter cells are born in a dome-like configuration (Fig. 5.1, 20.0h) which occupies a small area in two-dimensions. The sisters soon flatten out (Fig. 5.1, 21.5h) and begin to grow. At the end of the cell cycle (Fig. 5.1, 29.5-30.0h), the cell is seen to transform from an irregular, and often amorphous, shape into a nearly perfect circle, a side view reveals that the cell contracts laterally and rises vertically to become dome-like, similar to how it was born. The resulting division usually occurs laterally.

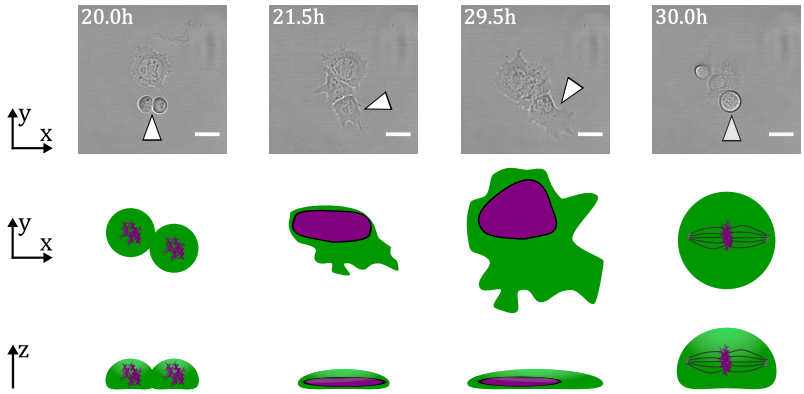


Fig. 5.1 Morphology Over the Cell-Cycle. The top row shows bright-field images for the time-course of a cell over its life (indicated by the white arrow), viewed from above. The middle row shows a parallel cartoon. The bottom row shows a cartoon of the cell profile at the same stages. Cells exhibit a regular seasonality over the cell-cycle. Sister cells are born in the dome-like configuration. The cells then flatten and progress through the cell-cycle, increasing in size over time. Before division, cells round up into a dome-like shape. (scale bar 20 μm)

The seasonality over one cell-cycle as described in Fig. 5.1 can be seen when the cell area quantified. Fig. 5.2 shows representative lineages of cell area for 2i ES cells and N2B27 cells. The cell area is initially small, corresponding to when the cell is dome-like. A sudden increase is seen when the cells flatten, at which point the rate of increase plateaus while the cells gradually becomes larger over time. Just before division, a sharp decrease in area coincides with the rounding up of the cell. Qualitative inspection of the replicates show the trend of a decrease in cell area over generations.

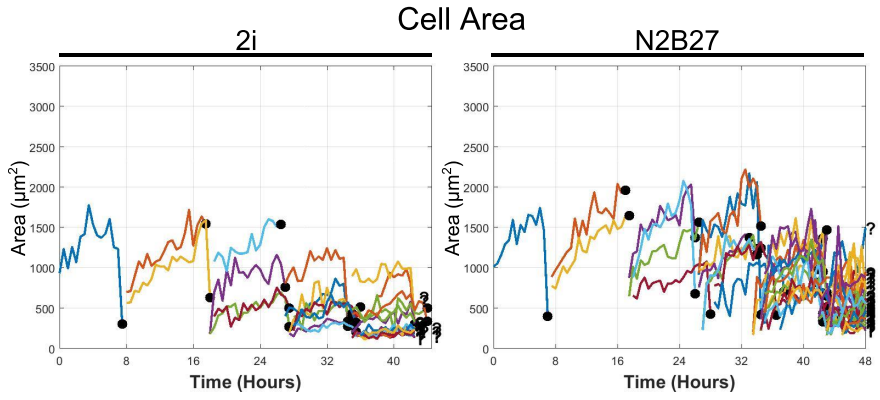


Fig. 5.2 Representative Time-Courses for Cell Area (μm^2). Cells are born dome-like with a small area. The area suddenly increases when they first flatten out which is followed by gradual growth. A sharp decrease in cell area is seen prior to division when cells "round up". Each line represents a cell and colors are arbitrary. Circles indicate division and '?' indicates that the fate of the cell is unknown.

The mean cell area decreases over subsequent generations (Fig. 5.3). Two-way ANOVA shows that this trend is significant for both replicates (p-value < 0.0001). As the number of cells in a colony increases, the area a cell occupies decreases. This is true for both 2i ES cells and exiting N2B27 cells. The areas of 2i ES cells tend to be smaller than N2B27 cells. Especially at later generations when N2B27 cells have exited (p-value < 0.0001). These data are summarized in Table 5.1.

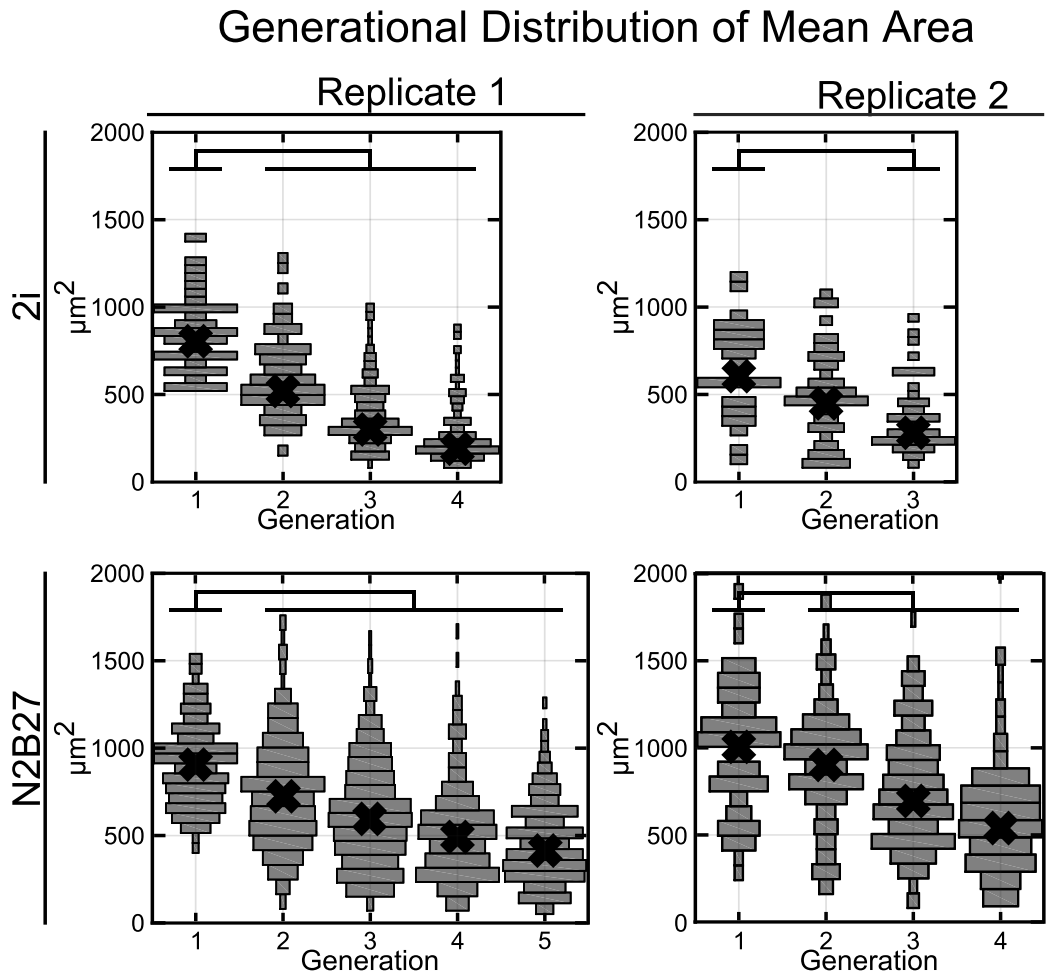


Fig. 5.3 Generational Distribution of Mean Cell Area. Aztec plots for the mean cell area (μm^2) are show for 2i ES cells and N2B27 cells. Two-way ANOVA shows that there is a decrease in cell area across generations for all conditions (p-value < 0.0001). This behavior is seen across replicates for both 2i and N2B27. However, 2i ES cells are consistently smaller than exiting N2B27 cells (p-value < 0.0001).

Table 5.1 Summary Statistics for Mean Cell Area. 2i ES cells and N2B27 cells are broken down by generations and biological replicates, sub-rows. The mean, standard deviation, number of cells ($\mathbf{N}_{cell}$), and number of clones ($\mathbf{N}_{clones}$) are shown for each.

Mean cell area (μm^2) by generation								
	2i				**N2B27**			
Generation	**Mean**	**SD**	$\mathbf{N}_{cell}$	$\mathbf{N}_{clone}$	**Mean**	**SD**	$\mathbf{N}_{cell}$	$\mathbf{N}_{clone}$
1	862.9	218.3	34	17	916.4	243.5	126	63
	643.6	259.3	30	15	1006.0	346.9	70	35
2	608.8	230.0	66	17	759.7	311.1	248	63
	500.6	264.1	44	12	908.2	345.3	122	32
3	402.8	202.9	122	17	632.6	301.0	360	52
	377.3	203.2	42	6	762.1	326.6	182	26
4	294.9	185.2	120	13	532.9	280.3	332	32
	–	–	–	–	611.3	313.1	130	14
5	–	–	–	–	462.5	235.2	204	14
	–	–	–	–	–	–	–	–

5.2.2 Motility Increases With the Onset of Exit

ES cells exhibit low motility, traveling short distances from where they were born (Fig. 5.4, top left). 2i ES cells are most active when they are born, and when they divide, remaining sedentary in between (Fig. 5.4, top right). On the other hand, when ES cells begin to down-regulate Rex1-GFPd2, indicating exit from naïve pluripotency, they are seen to travel greater distances (Fig. 5.4, bottom left). In early generations, cell movement is restricted to birth and divisions (Fig. 5.4, bottom right), but in later generations cells become motile throughout their lifespan. In videos, these colonies appear to spring to life, often moving *en masse* while the occasional cell is jettisoned from the periphery. The jettisoned cells often appear to be looking for neighbors with long cytoplasmic projections.

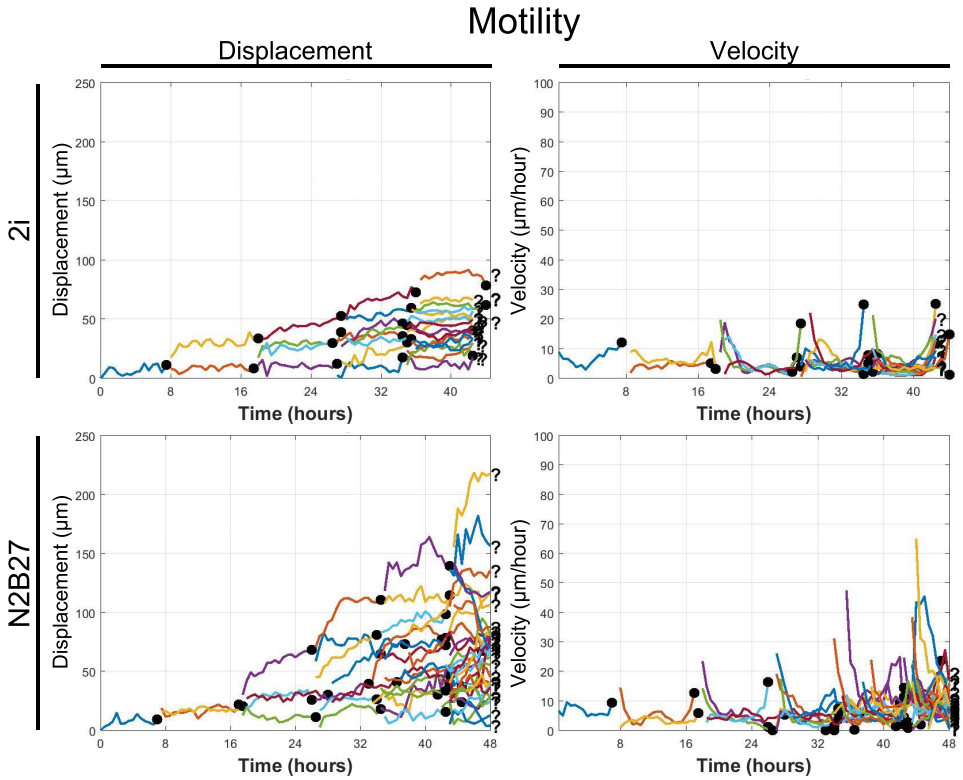


Fig. 5.4 Representatives Time-Courses for Cell Displacement and Cell Speed. The Time-course of displacement (μm) and velocity (μm/hour) for 2i ES cells and N2B27 Cells. 2i ES cells are qualitatively sedentary (top). They show short displacements (top left) and are only motile at birth and division (top right). Exiting N2B27 cells exhibit comparatively greater displacement (bottom left) and later generations are motile throughout the cell-cycle (bottom right). Each line represents a cell and colors are arbitrary. Circles indicate division and '?' indicates that the fate of the cell is unknown.

The average speed was used to quantify the motility of 2i ES cells and N2B27 cells (Fig. 5.5). Two-way ANOVA shows that there is decrease in the motility of 2i ES cells, meaning that as the conlony becomes larger, cells become less motile (p-value < 0.0001). This is not the case for the motility of exiting N2B27 cells. The longer ES cells remain in N2B27 (i.e. the further they move away from naïve pluripotency), the more motile they become. Two-way ANOVA shows a significant increase in the average speed of exiting N2B27 cells over subsequent generations (p-value < 0.0001). The initial few generations remain at a low level of motility, much like the 2i ES cells, but once exit begins, cell motility begins to increase. These data are summarized in Tables 5.2.

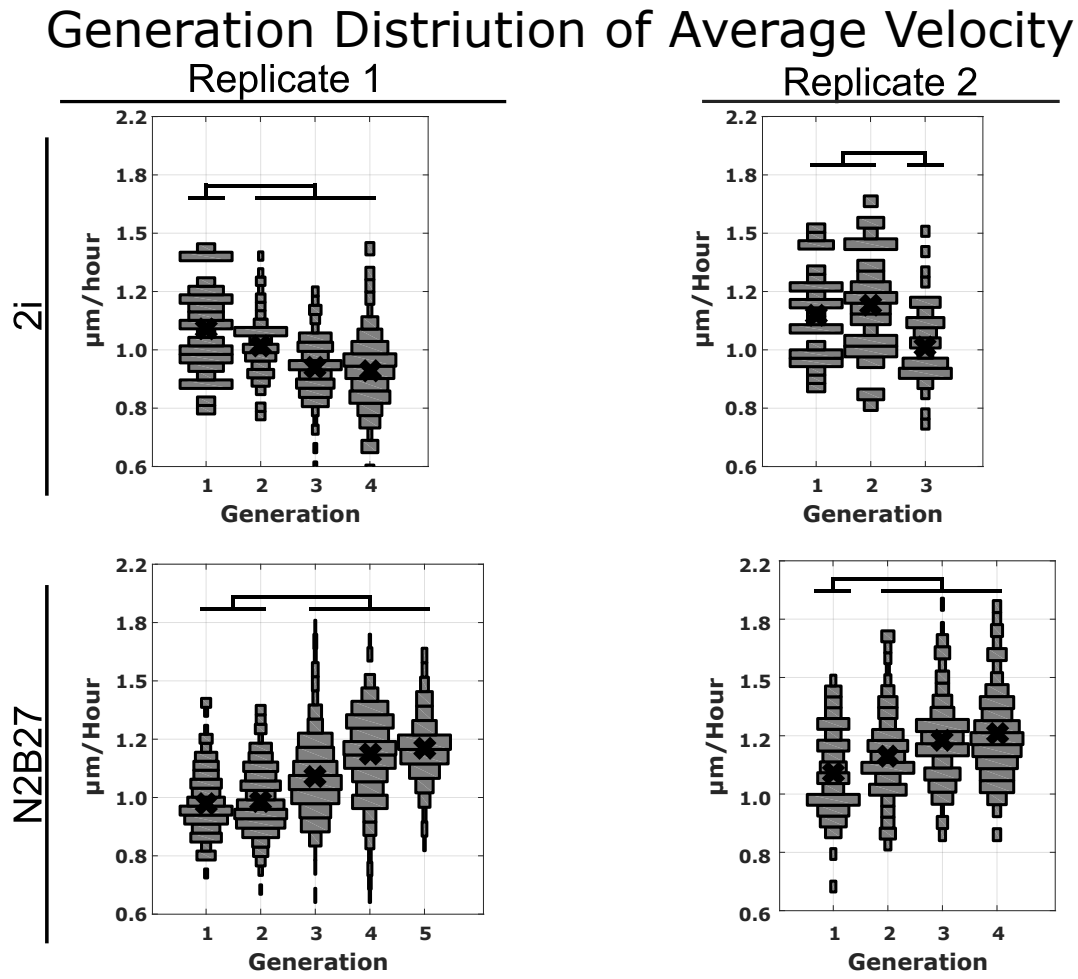


Fig. 5.5 Generational Distributions of Average Cell Speed. Aztec plots for average cell speed (μm/hour). Two-way ANOVA shows that, over generations, there is decrease in average velocity of 2i ES cells (p-value < 0.0001) but an increase is seen in N2B27 cells (p-value < 0.0001). Holm-Sidak post-test for exiting N2B27 cells shows a significant increase by the third generation for replicate 1 and the second generation for replicate 2.

Table 5.2 Summary Statistics for Average Cell Speed. 2i ES cells and N2B27 cells are broken down by generation and biological replicates, sub-rows. The mean, standard deviation, number of cells ($\mathbf{N}_{cell}$), and number of clones ($\mathbf{N}_{clones}$) are shown for each.

Average cell speed (μm/hour) by generation								
	2i				**N2B27**			
Generation	**Mean**	**SD**	$\mathbf{N}_{cell}$	$\mathbf{N}_{clone}$	**Mean**	**SD**	$\mathbf{N}_{cell}$	$\mathbf{N}_{clone}$
1	1.076	0.090	34	17	0.991	0.072	126	63
	1.139	0.093	28	15	1.092	0.092	70	35
2	1.012	0.068	66	17	0.999	0.074	244	63
	1.182	0.106	39	12	1.167	0.010	122	32
3	0.932	0.065	114	17	1.099	0.086	348	52
	1.028	0.081	42	6	1.229	0.095	177	26
4	0.915	0.095	91	13	1.164	0.090	294	32
	–	–	–	–	1.259	0.101	119	14
5	–	–	–	–	1.203	0.079	153	14
	–	–	–	–	–	–	–	–

5.2.3 Summary

A distinct morphological difference exists between 2i ES cell colonies and N2B27 cell colonies. As 2i ES cell colonies grow in number the area each cell occupies becomes ever smaller. They remain sedentary, only moving during the flattening after birth and at rounding upon divisions. N2B27 cells exhibit a similar pattern of cell area decrease with the increase in colony size, however they are consistently larger than 2i ES cells. Unlike 2i ES cells, N2B27 cells exhibit a higher motility in later generations, where cells exhibit motility throughout the cell-cycle and not only during birth and division.

5.3 The Relationship Between the Cell-Cycle and the Exit from Naïve Pluripotency

The cell-cycle has been implicated in several aspects of development and stem cell potency (Pauklin and Vallier, 2013, 2014; Snow, 1977; ter Huurne et al., 2017). Different culture systems have recently been shown to use different mechanisms of cell-cycle control, including

a lengthened G1 phase in 2i (ter Huurne et al., 2017). This is of particular interest when considering the removal of Mek inhibitor PD03, which blocks the Fgf signaling pathway, shown to be an important driver of differentiation in the early embryo and in stem cells (see Introduction, Fig. 1.6) (Burdon et al., 1999a; Lanner and Rossant, 2010). Not only does this act to prevent differentiation, but it also disrupts cell proliferation (Lanner and Rossant, 2010). So what happens to the cell cycle once 2i is withdrawn and Fgf signaling is allowed to resume uninterrupted? Long-term single-cell imaging has been used to quantify the individual cell-cycle durations for 2i ES cells and N2B27. This data is used to cross-validate the doubling times observed in Ch. 3.

The cell-cycle has also be implicated in lineage specification (Pauklin and Vallier, 2013, 2014). Here I use a proxy for cell-cycle phase to examine the relationship between the time of signal withdrawal and the phase of the cell-cycle. Finally, studies have shown correlation between sister cells in different ES cell conditions and in other cell types (Cannon et al., 2015; Sandler et al., 2015). This section concludes by examining the intergenerational and intragenerational correlation of cell-cycle duration for 2i ES cells and exiting N2B27 cells.

5.3.1 The Cell-Cycle Accelerates During Exit from Naïve Pluripotency

PD03 is a Mek inhibitor that blocks the pro-differentiation signals of Fgf. It is logical to think that cells with an inhibited response to a mitogen would cycle more slowly than those that do not. This behavior was seen in the serum based cultures in Sec. 3.4.2, where the doubling time of LPS ES cells (15 ± 1.6 hours) was longer than serum cells (10 ± 0.7 hours). However, in Sec. 3.3, I did not observe a difference between the doubling times of 2i ES cells and N2B27 cells, which both doubled their population every ≈ 12 hours. However, doubling time is a compound metric that combines the frequency of cell division, cell senescence, and cell death. Long-term single-cell tracking makes it possible to break down doubling time into its individual components. The individual cell-cycle duration was considered for 2i ES cells and N2B27 cells on a population basis and on a generational basis.

5.3.1.1 The Cell-Cycle of N2B27 Cells Accelerates, while the Cell-Cycle of 2i ES Cells Remains Constant

In the first instance, all cells that were observed were pooled together to generate the distribution of cell-cycle duration for the two culture conditions (Fig. 5.6). Both biological replicates show that the cell-cycle duration of 2i ES cells is longer than the cell-cycle duration of N2B27 cells (p-value < 0.0001). These data are summarized in Table 5.3.

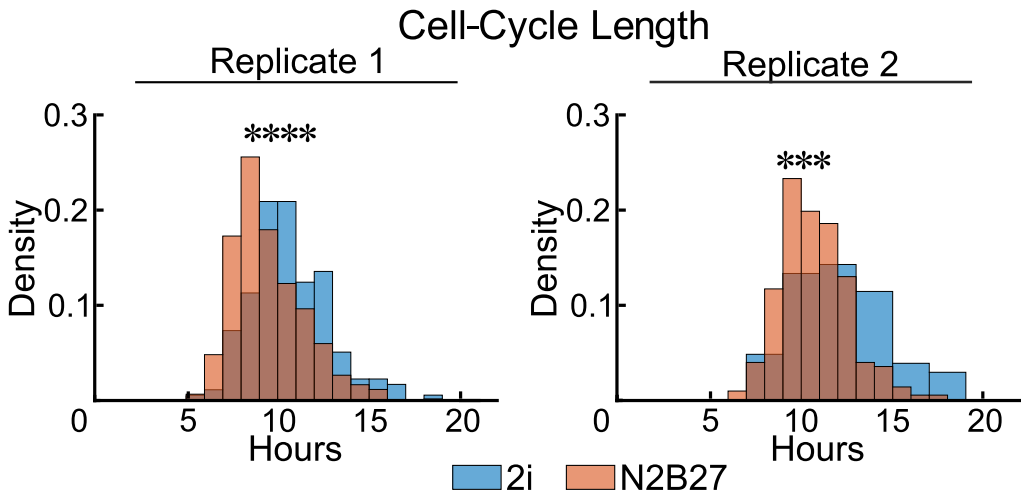


Fig. 5.6 Distribution of Cell-Cycle Duration for Cells in 2i and N2B27. 2i ES cells are shown in blue and N2B27 cells are shown in red. Two-way ANOVA shows, that for both biological replicates, N2B27 cells cycle more quickly than 2i ES cells (p-value < 0.0001).

Qualitative analysis of the images showed that cells began to divide more rapidly later in the time-course. Consequently, the distribution of cell-cycle durations was quantified for individual generations (Fig. 5.7). Two-way ANOVA shows that the cell-cycle duration of 2i ES cells remains constant over subsequent generations (p-value = 0.8354). On the other hand, N2B27 cells show a clear trend of an ever decreasing cell-cycle length, where all generations are significantly shorter than the first generations (p-value < 0.0001). These data are summarized in Table 5.3.

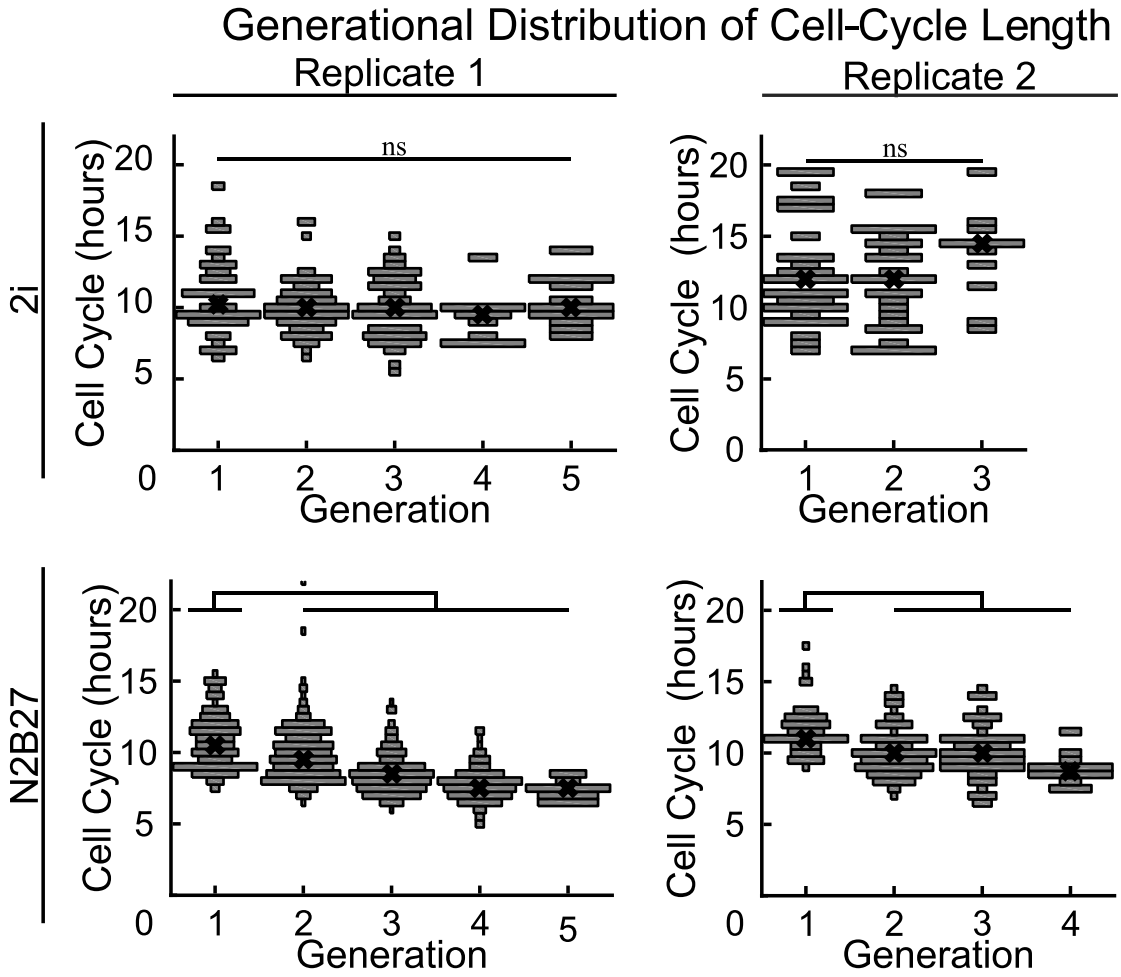


Fig. 5.7 Generational Distributions of Cell-Cycle Duration for Cells in 2i and N2B27. Cell cycle is further broken down by generations by Aztec plots for cell-cycle duration of 2i ES cells and N2B27. Two-way ANOVA shows that the cell-cycle duration of 2i ES cells remains constant over subsequent generations (p-value = 0.8354). However, there is a shortening of cell-cycle duration for N2B27 cells starting after the first generation (p-value < 0.0001).

Table 5.3 Summary Statistics for Cell-Cycle Duration for Cells in 2i and N2B27. 2i ES cells and N2B27 cells are broken down by generation and biological replicates, sub-rows. The mean, standard deviation, number of cells ($\mathbf{N}_{cell}$), and number of clones ($\mathbf{N}_{clones}$) are shown for each.

Interculture cell-cycle duration by generation								
	2i				**N2B27**			
Generation	**Mean**	**SD**	$\mathbf{N}_{cell}$	$\mathbf{N}_{clone}$	**Mean**	**SD**	$\mathbf{N}_{cell}$	$\mathbf{N}_{clone}$
All	10.41	2.12	177	23	9.26	2.03	602	63
	11.97	2.69	53	18	10.41	1.83	233	32
1	10.82	2.65	44	23	10.80	1.95	124	63
	12.44	3.60	31	18	11.57	1.66	61	32
2	10.06	1.84	69	19	9.94	2.03	180	52
	12.67	4.18	29	11	10.12	1.68	92	26
3	10.05	2.13	60	13	8.56	1.43	171	33
	13.68	3.14	11	3	10.07	1.80	66	15
4	9.27	1.75	11	3	7.79	1.24	102	14
	–	–	–	–	8.86	1.03	14	5
5	10.38	1.72	12	1	7.34	0.66	22	4
	–	–	–	–	–	–	–	–

5.3.1.2 There is No Distinction Between the Cell-Cycle Duration of Rex1-GFPd2 Positive and Negative Cells

Since a significant difference in cell-cycle duration exists between 2i ES cells and N2B27 cells, it follows that there may exist a difference between the two sub-populations in N2B27: Rex1-GFPd2 positive cells, which are recoverable back to naïve pluripotency, and Rex1-GFPd2 negative cells, cells that have irreversibly exited. The analysis from the previous sections was conducted on these two sub-populations.

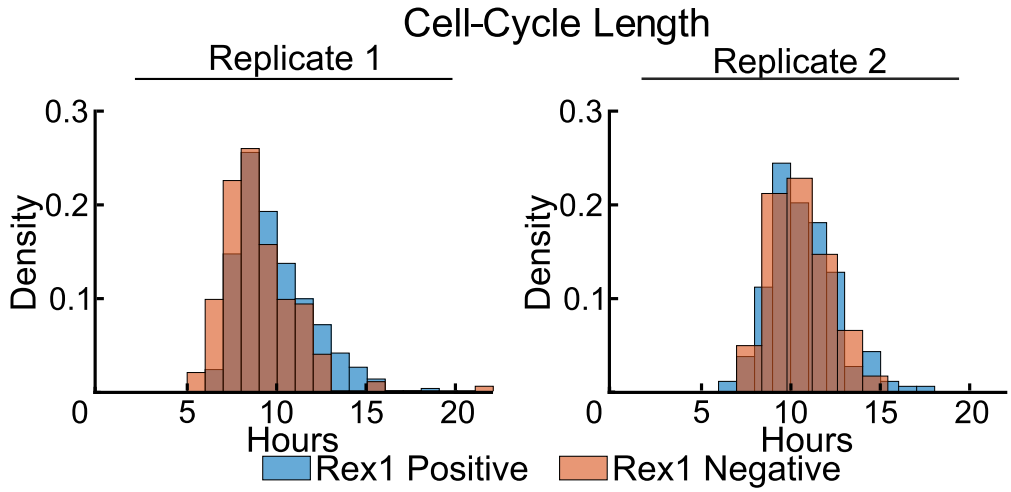


Fig. 5.8 Distribution of Cell-Cycle Duration for Rex1-GFPd2 Positive and Negative Cells in N2B27. Rex1-GFPd2 positive cells are shown in blue and negative cells are shown in red. No significant difference is observed between the two sub-populations in either biological replicate (p-value = 0.3546).

First, no difference between pooled Rex1-GFPd2 positive and negative cells was observed (Fig. 5.8, p-value = 0.3546). When, the individual generations were considered (Fig. 5.9), the trend of a decreasing cell-cycle was observed in both Rex1-GFPd2 positive and negative cells. As previously observed, for the case of all N2B27 cells, both sub-populations showed a significant difference over subsequent generations ($p < 0.0001$). But when the two sub-populations were compared against one another, two-way ANOVA showed no difference (p-value = 0.6849). These data are summarized in Table 5.4. I conclude that N2B27 cells proliferate faster, but there is no correlation between proliferation rate and exit, given that Rex1-GFPd2 Positive and Negative cells behave the same.

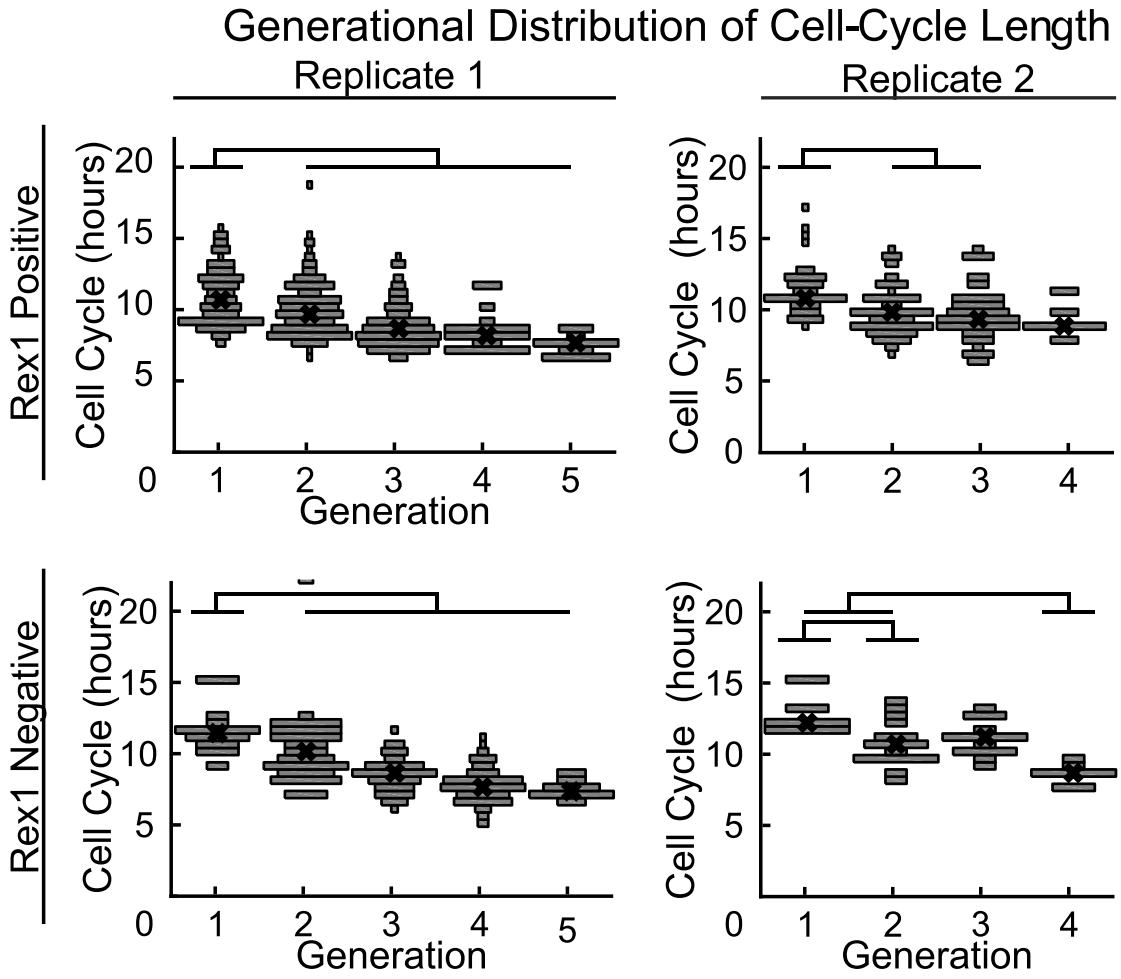


Fig. 5.9 Generational Distributions of Cell-Cycle Duration for Rex1-GFPd2 Positive and Negative Cells in N2B27. Cell Cycle duration is further broken down by generations by Aztec plots for cell-cycle duration of Rex1-GFPd2 positive and negative cells in N2B27. Two-way ANOVA shows that there is a significant decrease in the cell-cycle duration for both Rex1-GFPd2 positive and negative cells in N2B27 (p-value < 0.0001). However, there is no difference between the two sub-populations (p-value $= 0.6849$).

Table 5.4 Summary Statistics of Cell-Cycle Duration for Rex1-GFPd2 Positive and Negative Cells in N2B27. Cells are broken down by generation and biological replicates, sub-rows. The mean, standard deviation, number of cells ($\mathbf{N}_{cell}$), and number of clones ($\mathbf{N}_{clones}$) are shown for each.

Intraculture cell-cycle duration by generation								
	Rex1-GFPd2 Positive				**Rex1-GFPd2 Negative**			
Generation	**Mean**	**SD**	$\mathbf{N}_{cell}$	$\mathbf{N}_{clone}$	**Mean**	**SD**	$\mathbf{N}_{cell}$	$\mathbf{N}_{clone}$
All	9.56	1.99	397	57	8.69	1.98	205	40
	10.38	1.86	189	29	10.55	1.72	44	18
1	10.70	1.98	108	56	11.47	1.62	16	9
	11.47	1.67	55	29	12.50	1.34	6	4
2	9.88	1.90	135	42	10.12	2.38	45	21
	10.03	1.69	76	21	10.56	1.56	16	6
3	8.65	1.55	113	23	8.38	1.17	58	20
	9.85	1.88	52	14	10.89	1.18	14	7
4	8.25	1.25	24	6	7.65	1.20	78	13
	9.42	1.20	6	3	8.44	0.68	8	3
5	7.32	0.70	14	3	7.38	0.64	8	4
	–	–	–	–	–	–	–	–

5.3.2 Single-Cell Doubling Times Agree With The Population Data

To verify that the method of long-term single-cell imaging has had no ill effect on the culture, the growth curves of N2B27 cells and 2i ES cells compared against the growth curves determined from the bulk cultures. This was accomplished by first calculating the growth curves for each individual lineage (Fig. 5.10). The curves represent the number of traceable cells present at time t in a clone.

An issue arises, because some cells will become untraceable. Meaning the cells will either go outside the field of view or become unidentifiable for any number of reasons, including stacking upon one another, dense compaction, or losing a cell from one frame to the next. This unavoidable inability to trace cells leads to a drop in n, which cause the growth curves to become inaccurate. Notice how some of the growth curves in Fig. 5.10 suddenly drop. Therefore, instead of discarding valuable, time-intensive data after the first untraceable cell is incurred, an estimate of the total number of cells is made. This is done using the generational

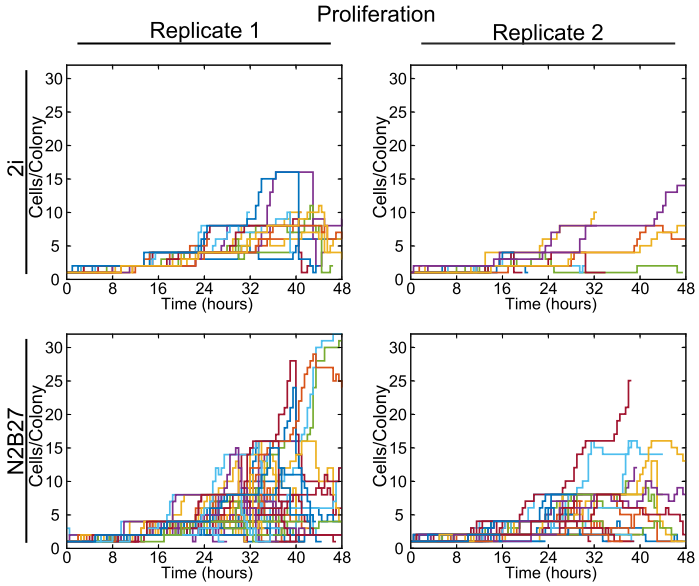


Fig. 5.10 Clonal growth curves for 2i ES cells (top row) and N2B27 cells (bottom row). Each line represents the growth curve of a single lineage. The curves initially take on the typical exponential growth pattern up to a certain point, and then exhibit a sharp down swing in cell number. This is due to a decrease in the the number of cells that are traceable within the colony, rather than cell death.

distributions of cell-cycle duration from the previous section. Growth curve calculations are described in detail in Sec. 2.3.6.

The raw growth cures are converted into the heat-maps in Fig. 5.11. These show the distribution of the number of cells per clone over time. Superimposed are two white curves. The lesser valued line is the mean of the raw data values. It can be seen in each case that, at a certain time t, the raw mean growth curve diverges from the familiar exponential growth curve; nor does is represent what is seen in the heat-map. This divergence is commensurate with a sudden drop in traceable cells. The greater valued curve is the calculated estimate. The estimated growth curves were then used as a comparison against the bulk-culture assays (Fig. 5.12). A red reference line is generated from the empirical values determined in Sec. 3.3.2. This is plotted with the standard deviation as the dashed red lines. The similarity of the curves show that the cell-cycle has been well maintained between the two assays.

5.3.3 Exit Kinetics are Unaffected by Cell-Cycle Timing

In human ES cells, there is evidence to suggest that specificaiton into particular lineages is restricted by the phase of the cell-cycle (Pauklin and Vallier, 2013, 2014). So it begs the question, is exit from the naïve state affected by cell-cycle timing? To answer this with the

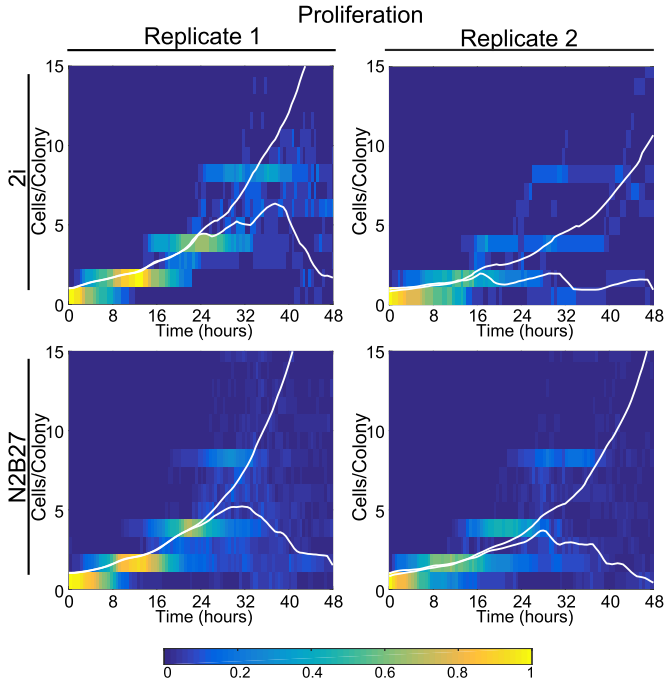


Fig. 5.11 Global Growth curves were used to estimate the missing data (see Sec. 3.3.2). The heat-maps show the raw growth curves, normalized by the total number of observations. The color bar show density of cells (count normalized by maximum count). Superimposed are two white lines represent the estimated (upper) and raw (lower) mean growth curves of all lineages.

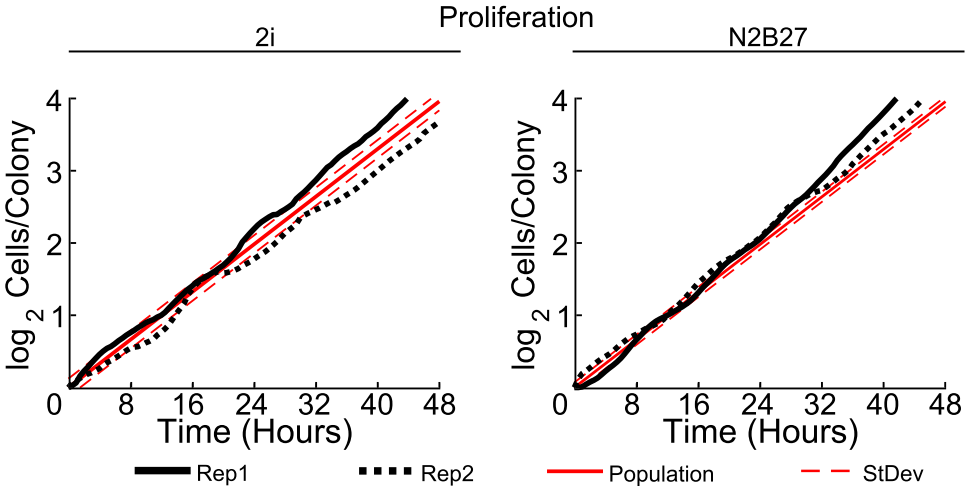


Fig. 5.12 Comparison of Single-Cell and Bulk-Culture Growth Curves. The long-term single-cell imaging growth curves were compared against the growth curves calculated from the bulk-culture data. The solid red references lines are the empirically calculated growth curves from the bulk-culture assays and the dashed red lines are the standard deviation. The black lines are the calculated growth curves from replicate one (solid line) and replicate 2 (dotted line) of the long-term single-cell imaging.

data at hand, the question must be rephrased. Does the phase of the cell-cycle at which self-renewal signals are withdrawn affect the kinetics of Rex1-GFPd2 down-regulation?

The time of first division was used as a proxy for the phase of the cell-cycle. For example, if the initial cell was witnessed to divide early in the video, it stands to reason that 2i was withdrawn late in the cell-cycle. Conversely, if the initial cell was seen to divide later in the video, it can be argued that 2i would have been withdrawn early in the cell-cycle. Fig. 5.13 shows the trajectories of Rex1-GFPd2 for individual branch-traces ordered by the times divisions occur (explanation in Sec. 2.3.5.). Both culture conditions show the seasonality of total Rex1-GFPd2 over each cell-cycle. As the cells grow, the signal intensity increases. Examining the trajectories of N2B27 cells shows a dampening and eventual absence of the fluorescence signal.

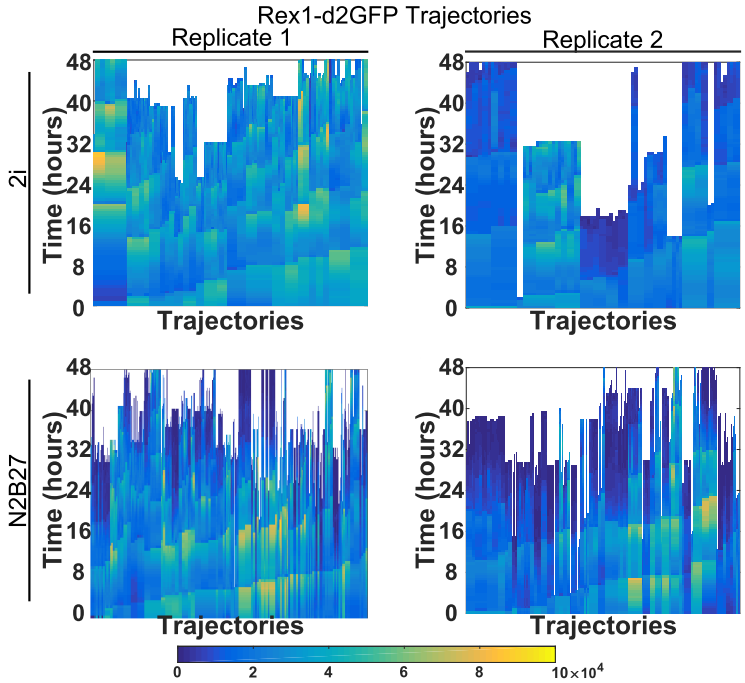


Fig. 5.13 Rex1-GFPd2 branch-trace trajectories were ordered by time of divisions to explore the relationship between expression levels and times of divison. The color bar indicates the value of total cellular Rex1-GFPd2.

It is not immediately apparent that any correlation exists between time of first division, the proxy for phase of the cell-cycle, and the time of exit. So to determine this relationship, these two attributes were quantified in two ways (Fig. 5.14). First, the correlation was shown on the basis of the clone (top row of Fig. 5.14), using the time at which the first cell of a lineage exits, as determined by thresholding. This shows no significant correlation (p-value > 0.05). The second analysis was performed for the time of exit for the individual branch

traces (bottom row of Fig. 5.14). Although this analysis did yield a significant correlation, the correlation coefficients are considerably low. Therefore there is no correlation between exit kinetics and the time in the cell-cycle at which self-renewal signals were withdrawn.

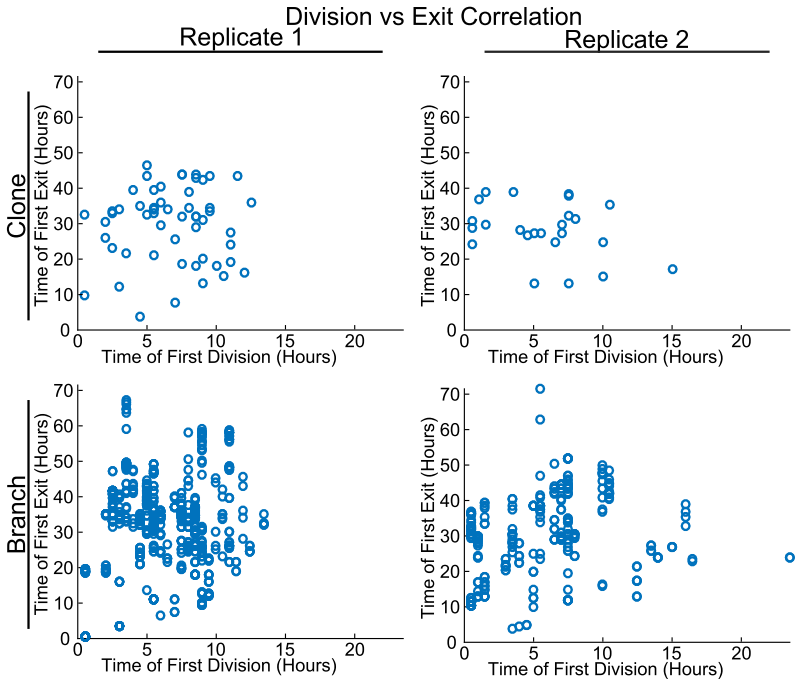


Fig. 5.14 The effect of signal withdrawal timing, in the the cell-cycle, on Rex1-GFPd2 down-regulation was examined by looking at the correlation between the time of first division and the time of first exit. This was considered on the basis of clones (top) and the basis of branch traces (bottom), however no strong correlation is apparent. Summary statistics are found in Table 5.5.

Table 5.5 Summary of correlation between time of exit and time of factor withdrawal. No strong correlation appears when this is considered for the time of first exit within a lineage, nor on the basis of individual branch traces. Pearson's-ρ is shown with the corresponding p-value.

Correlation of Cell-Cycle and Exit				
	Replicate 1		**Replicate 2**	
Basis	ρ	**p-value**	ρ	**p-value**
Clone	0.0441	0.7492	-0.3274	0.1183
Branch Trace	0.1765	< 0.0001	0.30457	<0.0001

5.3.4 The Cell-Cycle is Maintained Within Generations but Not Across Generations

The heritability of cell-cycle duration has been investigated for ES cells in different conditions and for other cell types (Cannon et al., 2015; Sandler et al., 2015), where high levels of correlations were shown to exist within generations (sisters), but not across generations (mothers and daughters). In this section my aim is to determine the similarity in cell-cycle heritability between genealogical relations of 2i ES cells and N2B27 cells.

Fig. 5.15 is the intragenerational analysis of the cell-cycle, showing the relationship between sisters and first-cousins (methods in Sec. 4.5.2). In the first set of analyzes, the cell-cycle durations of sister cells of 2i ES cells were compared to determine if they were the same or different. The distribution for difference in cell-cycle duration between sisters was calculated and produced a distribution centered on 0 hours, indicating little to no difference in cell-cycle duration (top-left panel in Fig. 5.15). A paired t-Test was conducted and confirmed that there was no statistical difference between the cell-cycle duration of sisters cells in 2i (p-value > 0.05). The same analysis was repeated for first-cousins in 2i, and again no significant difference was observed (p-value > 0.05). These results are expected because it was seen in Sec. 5.3.1 that the cell-cycle of 2i ES cells does not change from generation to generation. Surprisingly, this is also the case for sister and first-cousin N2B27 cells. This is surprising because the cell-cycle of N2B27 cells accelerates over subsequent generations.

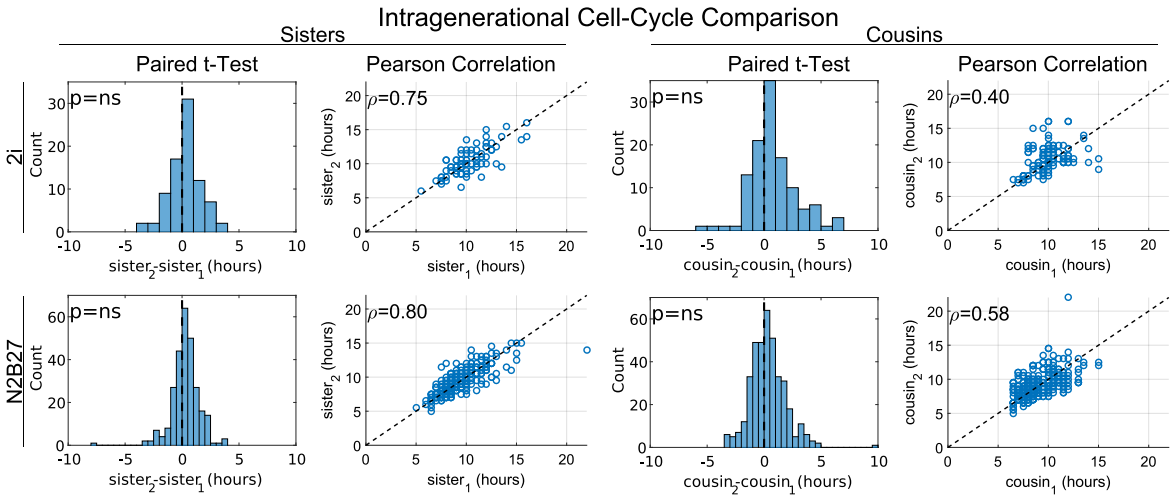


Fig. 5.15 Intragenerational genealogical analysis of cell-cycle duration shows that sisters are more correlated than cousins. Comparison between sisters and cousins, for both 2i ES cells and N2B27 cells. The paired t-Test shows the distribution of the difference between two cells with the associated p-value (p). The Pearson correlation shows a scatter plot of the paired cells with a reference line (black dashed line, slope = 1, intercept = 0) and the correlation coefficient (ρ). ns = not significant. Biological Replicate 1 is shown here, with all data summarized in Table 5.6

In a second set of analyzes, the Pearson correlation was used to determine the similarity between sisters and first-cousins (scatter-plots in Fig. 5.15). In the case of sister 2i ES cells, a high and significant correlation was observed ($\rho = 0.75$ and p-value < 0.0001). The correlation is clearly seen in the scatter plot in Fig. 5.15, where the cell-cycle duration or sisters are plotted against one another. Interestingly, both sisters and first-cousins in N2B27 showed a higher correlation coefficient than their 2i counterparts, indicating that during exit, cell-cycle durations are more coordinated than in the pluripotent state. These data are summarized in Table 5.6.

Similar comparsisons were performed for intergenerational relationships as controls (Fig. 5.16). The t-Tests show significance between both mothers & daughters and grandmothers & granddaughters. This is the case for both 2i ES cells and N2B27 cells. This is expected for N2B27 cells as the cell-cycle shortens over progressive generations. The low levels of intergenerational correlation and significance in 2i is in line with previous studies performed in different ES cell cultures and with different cell lines (Cannon et al., 2015; Sandler et al., 2015). These data are also summarized in Table 5.6.

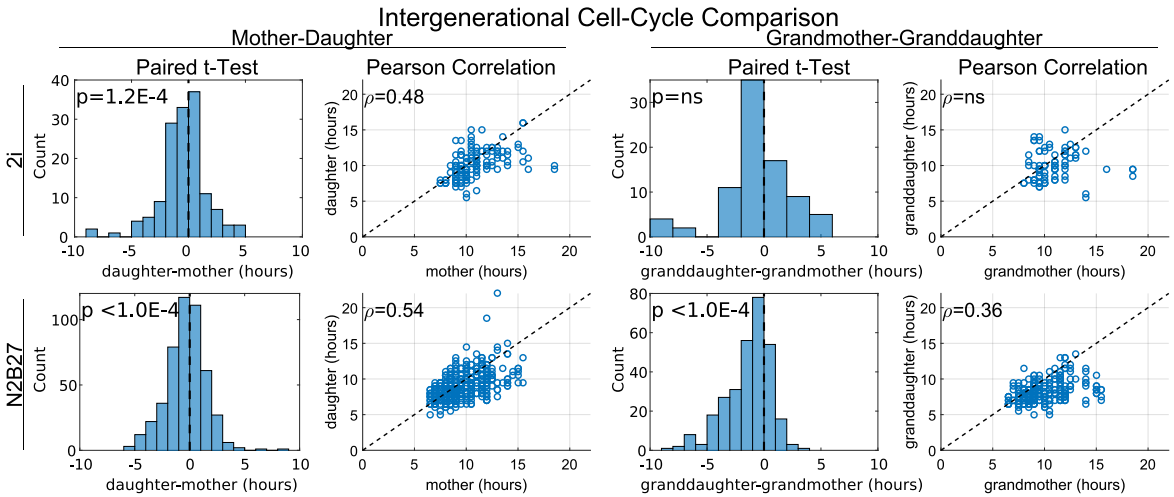


Fig. 5.16 Intergenerational genealogical analysis of cell-cycle duration show that correlation between generations is less than within generations. Comparison between mothers & daughters and grandmothers & granddaughters, for both 2i ES cells and N2B27 cells. The paired t-Test shows the distribution of the difference between two cells with the associated p-value (p). The Pearson correlation shows a scatter plot of the paired cells with a reference line (black ashed line, slope = 1, intercept = 0) and the correlation coefficient (ρ). ns = not significant. Biological Replicate 1 is shown here, with all data summarized in Table 5.6

Table 5.6 Summary of cell-cycle duration analysis for 2i ES cells and N2B27 cells. (t-test significance with p-value, Pearson correlation with p-value, number of cells, and number of clones.)

Genealogical Analysis of Cell-Cycle Duration												
	2i						**N2B27**					
	t-Test	p-value	ρ	p-value	N_{cell}	N_{clone}	t-Test	p-value	ρ	p-value	N_{cell}	N_{clone}
ss	ns	0.72	0.76	<1E-4	82	17	ns	0.51	0.81	<1E-4	271	62
	ns	0.65	0.61	1.3E-3	25	10	ns	0.16	0.88	<1E-4	101	30
cc	ns	6.6E-3	0.40	<1E-4	115	16	ns	0.21	0.59	<1E-4	386	45
	ns	5.3E-3	0.53	ns	28	5	ns	4.0E-3	0.61	<1E-4	136	24
md	***	1.2E-4	0.48	<1E-4	144	17	****	<1E-4	0.55	<1E-4	478	52
	ns	0.59	0.30	ns	31	6	ns	1.5E-2	0.29	1.1E-4	171	26
gd	ns	5.8E-3	0.05	ns	83	13	****	<1E-4	0.37	<1E-4	298	33
	ns	0.53	0.84	2.3E-3	10	2	ns	6.9E-3	0.34	1.9E-3	80	15

5.3.5 Summary

Long-term single-cell tracking has allowed for the examination of cell-cycle duration for individual cells. Here I determined that 2i ES cells have a longer cell-cycle than N2B27 cells. Furthermore, 2i ES cells maintain their cell-cycle duration over subsequent generations, whereas N2B27 cells exhibit progressively shorter cell-cycles over each generation. But no difference in cell-cycle duration was seen when the N2B27 cells were divided among those that had irreversibly exited (Rex1-GFPd2 negative) and those that had not (Rex1-GFPd2 positive). However, both sub-populations retain the behavior of generational acceleration in cell-cycle duration. These data were then used to estimate the growth curves for the different conditions. Comparison against the bulk-culture proliferation assays show these data match fairly well. Although the N2B27 cell growth curves were slightly steeper at the end. It would be interesting to determine if there is a difference in death rate between 2i and N2B27.

The effect of cell-cycle timing was considered in relation to the kinetics of exit from the naïve state. Ordering of branch-trace trajectories by time of divisions show a regular seasonality in Rex1-GFPd2 expression over the cell-cycle. However, correlation analysis between the time-of-exit and the time-of-first-division yields no correlation. Therefore, the kinetics of Rex1-GFPd2 down-regulation are unaffected by the point in the cell-cycle at which self-renewal signals are withdrawn.

Finally, examination of the genealogical relation of the cell-cycle between 2i ES cells and N2B27 cells reveals two features. First, it shows that ES cells are more highly correlated as they exit naïve pluripotency than when they are self-renewing. Second, that the closer two cells are in relation to one another, the more similar their cell-cycle will be. This is seen in the intragenerational and intergenerational analysis, showing that cells within a generation are highly correlated whereas cells across generations are not.

5.4 The Kinetics of Rex1-GFPd2 Down-Regulation

Kalkan et al., 2017, showed that upon the withdrawal of self-renewal signals, not all cells exit the naïve state immediately. Rather, they dwell in a reversible state for a period of time, where reapplication of self-renewal signals is sufficient for their recovery. However, without this reapplication these cells will continue to progress and eventually exit. This asynchronicity of exit from naïve pluripotency has been documented by Kalkan et al., 2017, and here in Ch. 3, but this has not been shown in real time as the exit occurs. So the actual kinetics of Rex1-GFPd2 down-regulation have not been observed and quantified. Long-term single-cell imaging has been used to quantify the real-time and shifted-time kinetics of Rex1-GFPd2 down-regulation.

The last section considers the symmetry of cell division. Habib et al., 2013, showed that asymmetric divisions can be forced *in vitro* by generating chemical inhomogeneities, such as the introduction of a Wnt covered bead. To date, no study has investigated the pattern of divisions during exit from naïve pluripotency. Here I determine the symmetry of unforced divisions as ES cells exit naïve pluripotency.

5.4.1 ES Cells Exit Asynchronously on the Single-Cell Level

Bulk-culture assays have shown that exit from naïve pluripotency occurs asynchronously (Ch. 3 & Kalkan et al., 2017). Long-term single-cell imaging now allows for the direct observation of asynchronously exiting lineages. Fig. 5.17 shows the time-course of total fluorescence for N2B27 cell lineages from both biological replicates. It can be seen that exit occurs at different times and in different generations. In contrast, 2i ES cells display a seasonality with the cell cycle yet remain Rex1-GFPd2 positive. Padovan-Merhar et al., 2015, attribute this seasonality to a transcriptional scalability with cell volume.

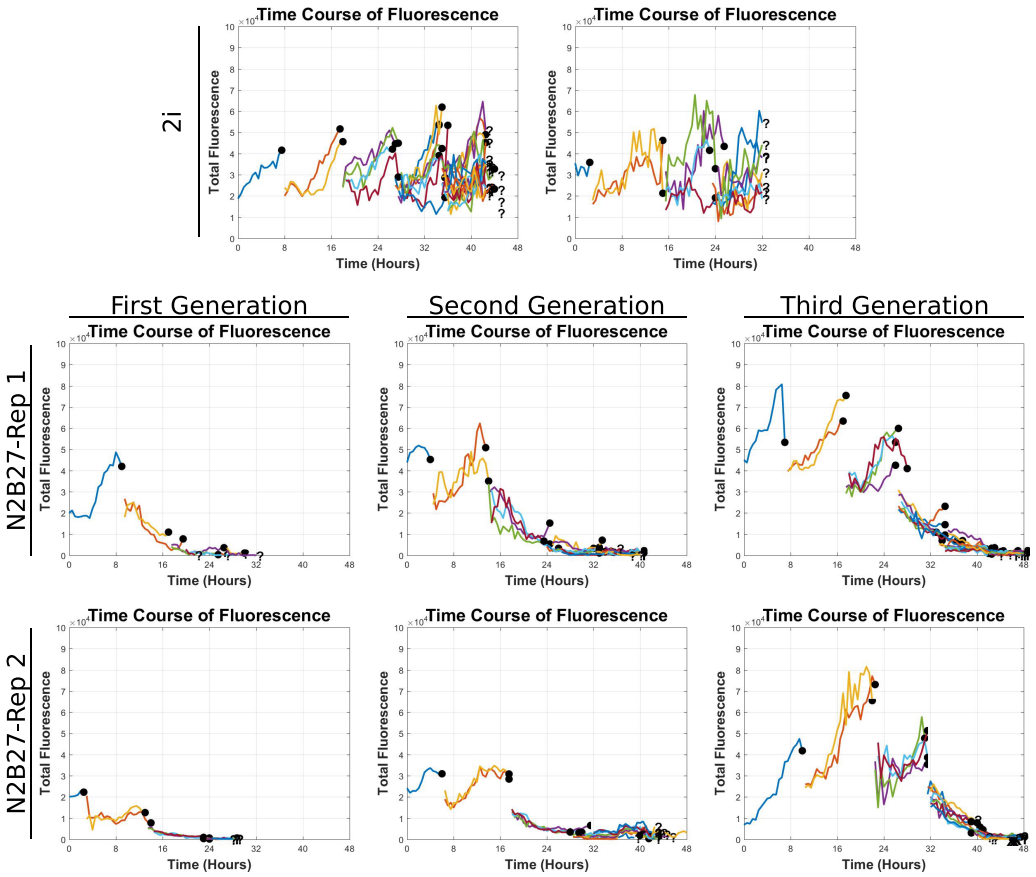


Fig. 5.17 A common trend of exponential decay appears within exiting lineages whose onset can be seen at different times and in different generations. Representative lineages for the time-course of total cellular fluorescence of Rex1-GFPd2. Each line represents a cell trajectory. Black circles indicate division and '?' indicate that the fate of a cell is unknown. The top row shows the expression pattern of 2i ES cells from each biological replicate. The bottom two rows show representative N2B27 cells down-regulating Rex1-GFPd2 in different generations.

All lineages of 2i ES cells and N2B27 cells we superimposed and shown as a heat-map in Fig. 5.18. These figures show the global distribution of Rex1-GFPd2 expression for all cells. The white lines show the estimated bounds of the mean global Rex1-GFPd2 expression level (method shown in Sec. 2.3.6). For 2i, the mean expression is relatively constant over time, the variability coming from the seasonality of expression level during the cell-cycle. Whereas the N2B27 cells show an onset of down-regulation that persists over many hours due to the asynchronicity of exit.

These data were converted to the proportion of cells that have exited, so they could be compared to the bulk-culture data. This was done by applying the determined threshold (see Methods Sec. 2.3.4) for when the cell trajectory is positive and negative (Fig. 5.19). This

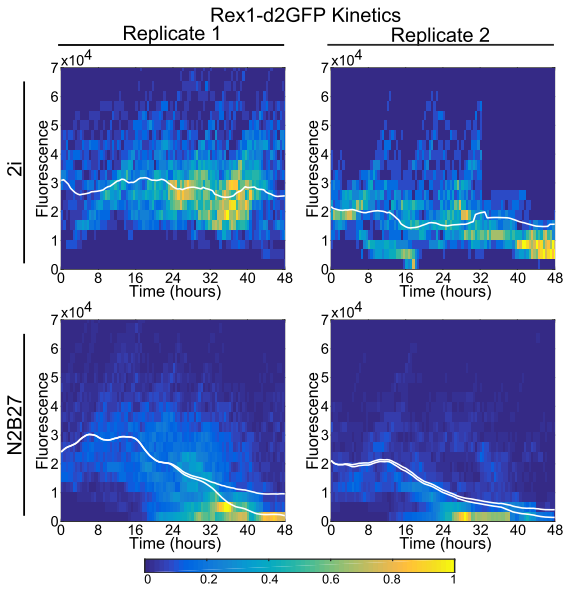


Fig. 5.18 When all lineages are viewed in real time, a trend of gradual down regulation can be seen. Heat-maps show the distribution of Rex1-GFPd2 expression levels. The colorbar is the density of cells. The white lines, the weak estimate (upper) and the strong estimate (lower), show the global mean Rex1-GFPd2 expression level.

was done with the Strong and Weak estimations for Rex1-GFPd2 expression (see Methods Sec. 2.3.6). The estimates can be seen to be comparable to the bulk-culture assay from Ch.3 indicating that the population of ES cells are exiting asynchronously.

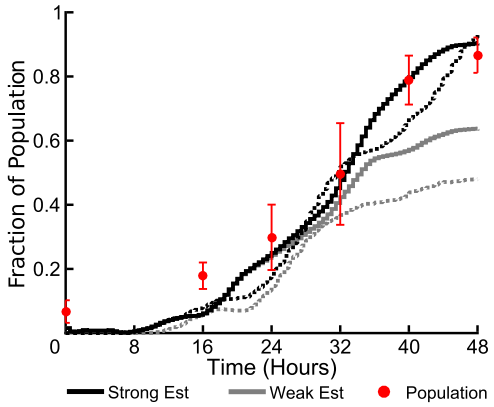


Fig. 5.19 The exit kinetics from the long-term single-cell tracking and the population-level are in agreement. Exit kinetics comparison for long-term single-cell imaging and bulk-culture flow cytometry (red points, mean ± SD). The black lines are the exit kinetics calculated using the Strong estimate and the gray lines are the exit kinetics calculated using the Weak Estimates (see Methods Sec. 2.3.6). These two curves provide upper and lower bounds for the actual percentage of cells that have exited the naïve state.

5.4.2 Rex1 is Regulated by a Switch Mechanism

Both the previous section, Sec. 5.4.1, and Ch. 3 have established that the exit from naïve pluripotency occurs asynchronously. However, the kinetics of Rex1-GFPd2 have not been established. For instance, it is not known if cells inactivate the Rex1-GFPd2 reporter with the same kinetics or with different kinetics. The individual lineages in Fig. 5.17 show a sudden onset of exponential decay. However, this behavior is blurred in global mean Rex1-GFPd2 expression levels due to the events occurring at different times. This asynchronicity can be artificially eliminated *in silico* by shifting the cell trajectories in time (method in Sec. 2.3.5). In the first instance, the simplest method was employed to align these trajectories. For each lineage, a new origin was set to the time at which the first cell of the clone exits. This new time zero is established with the on/off threshold that has been calculated (method in Sec. 2.3.4). When the time shifts are carried out, the raw trajectories tighten up and the exponential decay trend appears on the global level (Fig. 5.20).

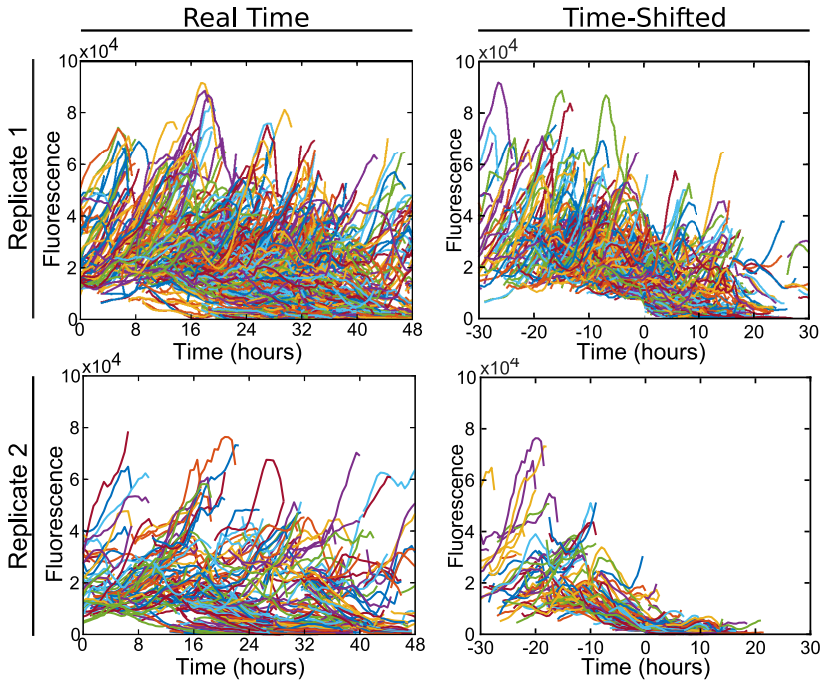


Fig. 5.20 To better resolve the common trend of exponential decay, the genetic trajectory for each lineage was shifted in time according to the time at which the first cell crossed the on/off threshold. All real-time lineages (left) are shifted in time (right) by the time the first cell crosses the on/off threshold.

Fig. 5.21 (left) shows the global distribution of all cell trajectories shifted in time. These figures are superimposed with the new Strong and Weak estimates of the global average Rex1-GFPd2 expression level. The trend that appears is a sudden onset of exponential decay,

implying that ES cells suddenly shut off Rex1-GFPd2. The the remaining signal would come from residual mRNA still being translated and from protein yet to decay. This is consistent with a switch mechanism, whereby Rex1-GFPd2 is rapidly and irreversibly switched off.

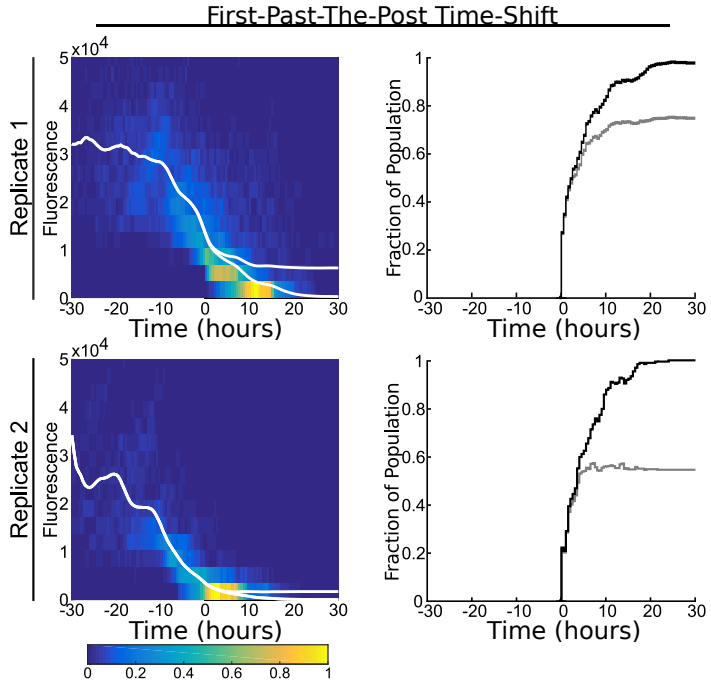


Fig. 5.21 The time-shifting genetic trajectories show a defined trend of exponential decay. Left, heat-maps show the distribution of Rex1-GFPd2 expression levels after genetic trajectory time-shifts. The colorbar is the density of cells. The white lines, the weak estimate (upper) and the strong estimate (lower), show the global mean Rex1-GFPd2 expression level. Left, The Strong estimate (black lines) and Weak estimate (gray lines) were used to calculate the exit kinetics of the time-shift. These two curves provide upper and lower bounds for time-shifted exit from the naïve state.

Using these data to calculate the proportion of cells that have exited, the exit kinetics can be seen to progress from a sigmoidal curve (Fig. 5.19) to a hockey stick curve (Fig. 5.21, right). This has implications for how synchronized cells are within a clone. For this alignment, the entire lineage has been shifted, but the individual trajectories have not been shifted relative to one another. So, this tight and rapid exit that is witnessed shows that there is a high level of intraclonal synchronicity.

5.4.2.1 Improving Time-Shifting of Genetic Trajectory Multiple Alignment

The current method used for time-shifting of genetic trajectory alignment of lineages to establish the global kinetics of Rex1-GFPd2 down-regulation kinetics can be improved upon

by using hierarchical clustering. Instead of shifting entire lineages, each branch-trace could be aligned. This would be accomplished by performing hierarchical clustering followed by minimization of the distance between curves and clusters. Fig. 5.22) shows an initial attempt at implementing hierarchical clustering. 16 clusters were generated using MATLAB's self organizing map from the neural network toolbox, however, a fully branched dendrogram would need to be constructed to properly conduct the multiple alignment for the curves. Nonetheless, they display a coherent and rapid reduction in Rex1-GFPd2 expression levels. These analyzes are still preliminary, but in the future it would be interesting to determine whether different exit kinetics exist or if the slightly dissimilar trends observed in Fig. 5.22 belong to a single family of Brownian paths.

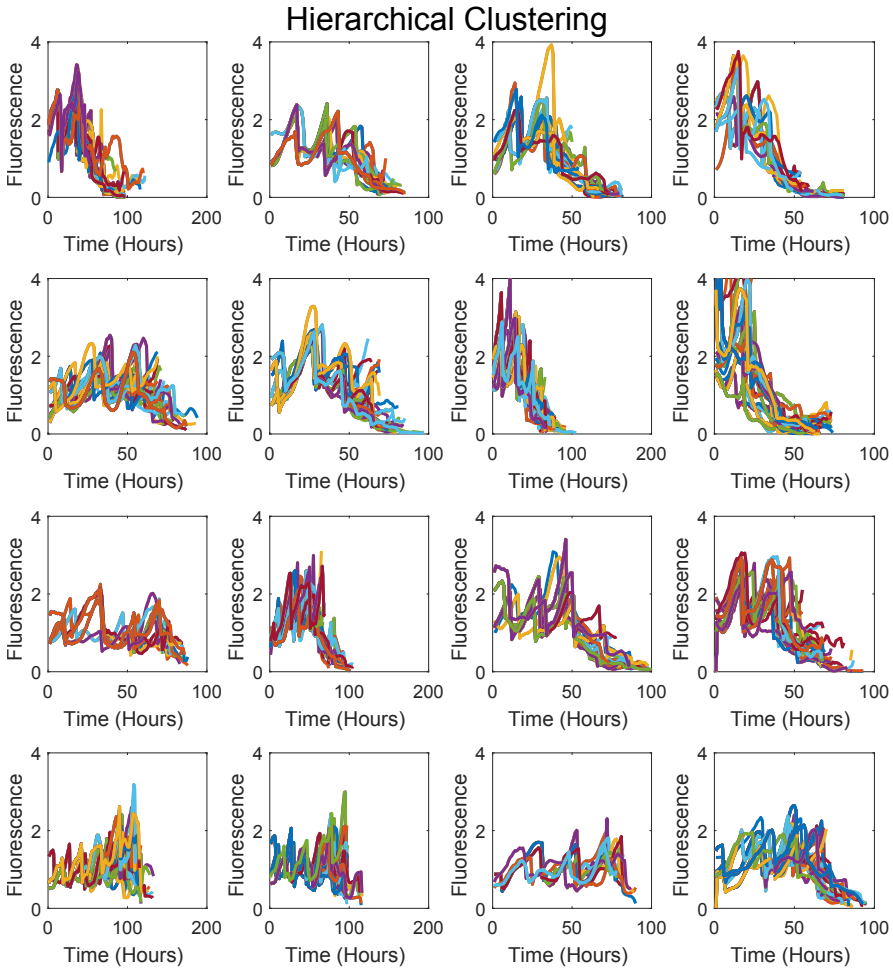


Fig. 5.22 Future work to improve time-shifting of genetic trajectories could incorporate hierarchical clustering techniques. Hierarchical clustering of branch-traces via a self-organizing map (MATLAB, Neural Network Toolbox). All branch-traces from replicate 1 have been sorted into 16 clusters. This preliminary clustering was done to show the need for improvement upon the current multiple alignment method.

5.4.3 ES Cells Exit via Symmetric Divisions

The two defining features of ES cells are self-renewal and the capacity to differentiate. At each division, as ES cell exit naïve pluripotency, the mother ES cell could produce two naïve daughters, two exited daughters, or one of each. Habib et al., 2013, shows that asymmetric divisions can be induced in ES cell with the introduction of an extrinsic chemical inhomogeneity, in that case a Wnt-coated bead. Brown et al., 2017, have show that between between 30-50% of divisions in differentiating human ES cells engage in asymmetric divisions after the first division. However, the previous section made note of high intraclonal synchronicity in the Rex1-GFPd2 signal, as shown in many lineages of Fig. 5.17. This section demonstrates that, as ES cells exit naïve pluripotency, divisions occur symmetrically.

There are several metrics by which divisions symmetry could be demonstrated (Fig. 5.23). These include the integrated fluorescence over the lifetime of the cell, mean lifetime fluorescence, the final fluorescence, and the rate of change in the fluorescence signal. The common feature of these attributes is that each is packaged with environmental inputs gathered over the lifetime of the cell. These environmental inputs have the potential to induce asymmetries over time. For example, if the two daughters land in different neighborhoods, the center of a colony versus the periphery of the colony. Over time, neighbors and paracrine signaling may have an effect on the genetic trajectory of the cells.

Thus to mitigate environmental effects, the initial Rex1-GFPd2 signal has been used to quantify cell division symmetry. The mean lifetime Rex1-GFPd2 signal is considered in parallel as an environmental control. Fig. 5.24 shows the generational distributions of initial Rex1-GFPd2 signal. The positive control, Rex1-GFPd2 cells in 2i, shows a relatively steady expression levels from generation to generation. Whereas the exiting cells, Rex-GFPd2 cells in N2B27, show the decline from positive to negative over several generations. The negative control, E14 IVC WT ES cells in 2i, do not contain the reporter so all that is seen is technical noise.

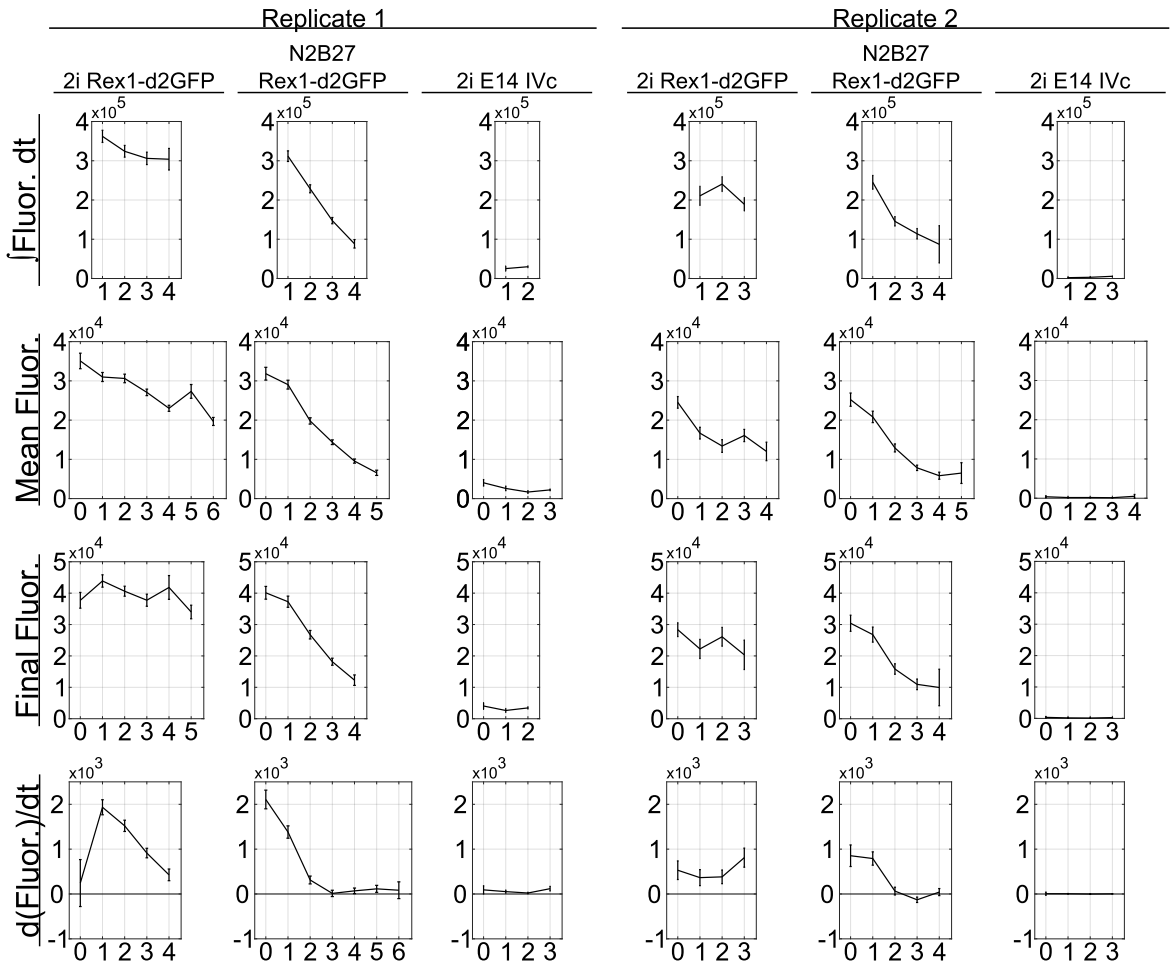


Fig. 5.23 Testing the symmetry of divisions based upon Rex1-GFPd2 levels could be performed via several different metrics, however, appropriate metrics need to be selected to mitigate confounding exogenous influences.Line graphs for the different metrics that could be used to investigate the symmetry of divisions symmetry. The positive control (2i Rex1-GFPd2), Exiting ES cells (N2B27 Rex1-GFPd2), and negative control (2i E14 IVC) are shown for each attribute. The line graphs show the mean±SEM for each generation.

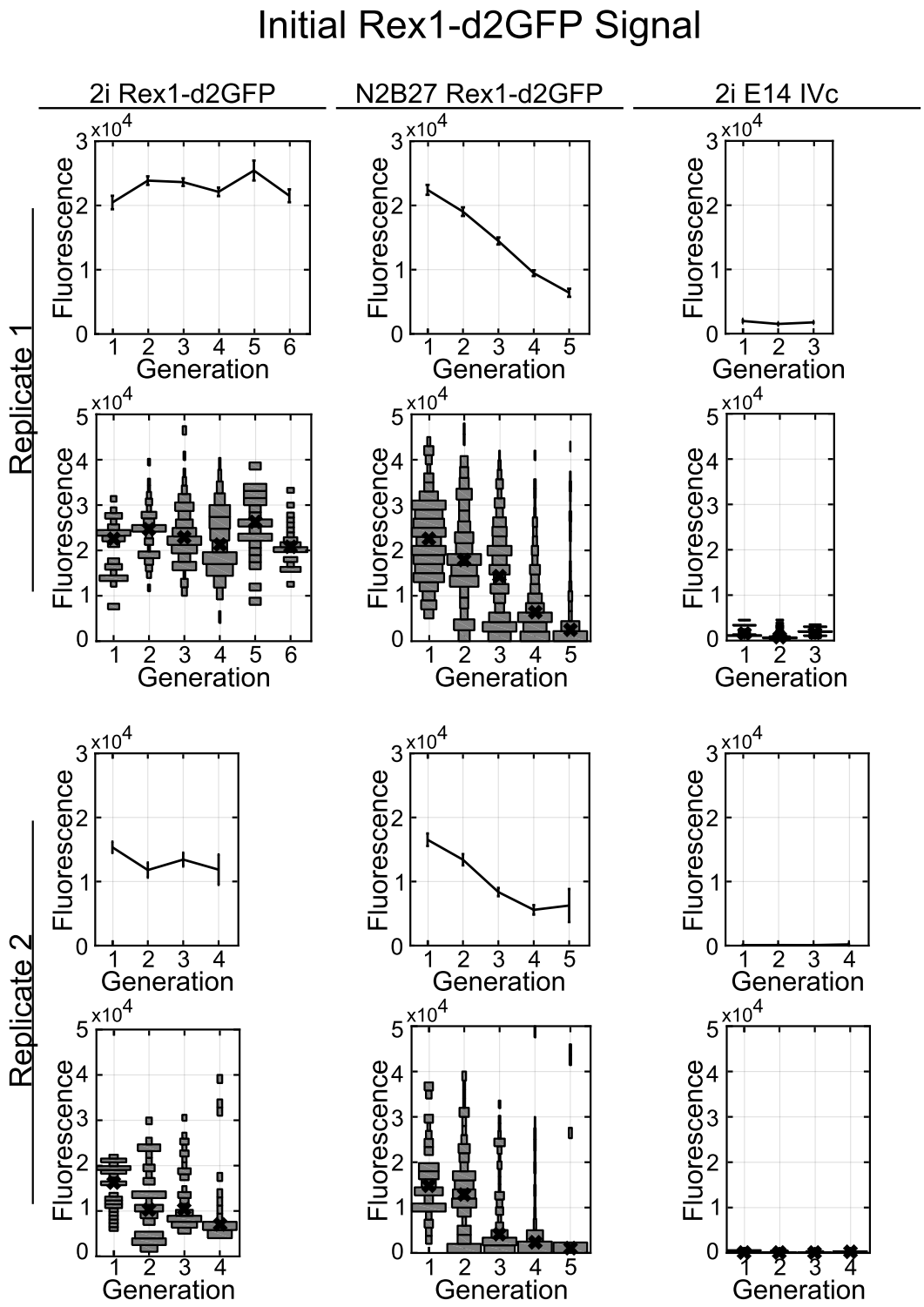


Fig. 5.24 The initial Rex1-GFPd2 level was selected to test the symmetry of cell division. Generational kinetics of the initial Rex1-GFPd2 level for the positive control (2i Rex1-GFPd2), Exiting ES cells (N2B27 Rex1-GFPd2), and negative control (2i E14 IVC). The initial fluorescence for the positive control remains fairly constant over generations while exiting cells down-regulate over generations. The negative control shows no change as the only signal comes from the technical noise in imaging. The line graphs show the mean and standard error of the mean and the Aztec plots show the generational distributions (black 'x' indicates the median).

Genealogical comparisons, similar to the cell-cycle comparisons in Sec. 5.3.4, were performed for the initial Rex1-GFPd2 signal (Fig. 5.25). In 2i, where all cells are pluripotent, the initial Rex1-GFPd2 signal is not significantly different in sisters. This is also the case for N2B27 cells, indicating symmetric inheritance of Rex1-GFPd2. Furthermore, correlation analysis show that N2B27 cells are more correlated than 2i ES cells (p-value < 0.05). Extending this analysis to cousins, replicate one shows a significant difference in 2i ES cells, while both replicates of N2B27 cells show no differences. Similarly, the correlation of N2B27 cells is greater than 2i ES cells. Again, as a control, mothers & daughters and grandmothers & granddaughters were compared and found to be significantly different for both 2i ES cells and N2B27 cells. This is more pronounced for N2B27 cells as expected, because differentiating cells down-regulate Rex1-GFPd2 over different generations. These data are summarized in Table 5.7.

A similar trend is seen for the mean Rex1-GFPd2 signal (Fig. 5.26 and Table 5.8). Over the life of the cells, sisters and first-cousins remain highly correlated as N2B27 cells exit. While mothers & daughters and grandmothers & granddaughters show a significant different in their means. Furthermore, the mean Rex1-GFPd2 signal shows higher levels of correlation than 2i ES cells (p-value < 0.05).

The three sets of genealogical analysis are shown together in 5.27. For the t-Test, sisters are the same in both 2i and N2B27, where first-cousins show no significance in N2B27 but show mixed results in 2i. Across generations, the trend is that all relationships are different. The exception is seen in the cell-cycle for 2i ES cells in replicate 2. In this case the correlation of mothers & daughters is not significant and the grandmother & granddaughter calculation is based upon only 2 clones.

The correlation data show high levels of correlation between the sisters and first-cousins of N2B27 cells. The level of correlation decreases with intergenerational comparisons and the distance of the relationship. 2i ES cells also show high correlation for sisters, but lower levels for first-cousins and intergenerational comparisons. These data present strong evidence for symmetric divisions during the process of exit from naïve pluripotency.

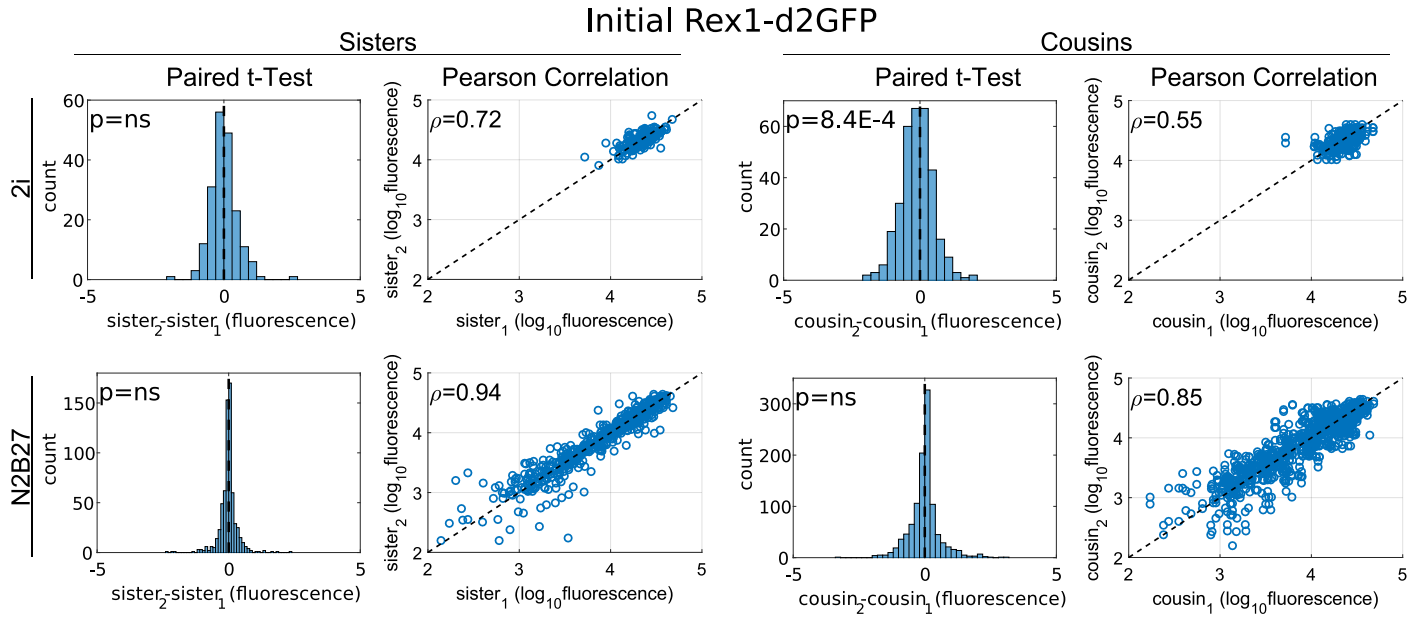


Fig. 5.25 Genealogical analysis of initial Rex1-GFPd2 signal. Comparison between sisters and cousins, for both 2i ES cells and N2B27 cells. The paired t-Test shows the distribution of the difference between two cells with the associated p-value (p). The Pearson correlation shows a log-log scatter plot of the paired cells with a reference line (black dashed line, slope = 1, intercept = 0) and the correlation coefficient (ρ). ns = not significant. Biological Replicate 1 is shown here, with all data summarized in Table 5.7

Table 5.7 Summary of initial Rex1-GFPd2 analysis for 2i ES cells and N2B27 cells. (t-test significance with p-value, Pearson correlation with p-value, number of cells, and number of clones.)

Genealogical Analysis of Initial Rex1-GFPd2												
	2i						**N2B27**					
	t-Test	p-value	ρ	p-value	N_{cell}	N_{clone}	t-Test	p-value	ρ	p-value	N_{cell}	N_{clone}
ss	ns	0.94	0.73	<1.0E-4	194	17	ns	0.46	0.95	<1.0E-4	660	63
	ns	3.5E-2	0.86	<1.0E-4	68	15	ns	0.43	0.97	<1.0E-4	265	35
cc	***	8.4E-4	0.56	<1.0E-4	328	17	ns	0.72	0.85	<1.0E-4	1080	62
	ns	2.0E-2	0.71	<1.0E-4	100	10	ns	0.71	0.86	<1.0E-4	400	30
md	**	1.3E-3	0.47	<1.0E-4	354	17	****	<1E-4	0.66	<1.0E-4	1194	63
	****	<1E-4	0.80	<1.0E-4	106	12	****	<1E-4	0.72	<1.0E-4	460	32
gd	ns	6.0E-3	0.29	<1.0E-4	288	17	****	<1E-4	0.18	<1.0E-4	946	52
	****	<1E-4	0.63	<1.0E-4	62	6	****	<1E-4	0.43	<1.0E-4	338	26

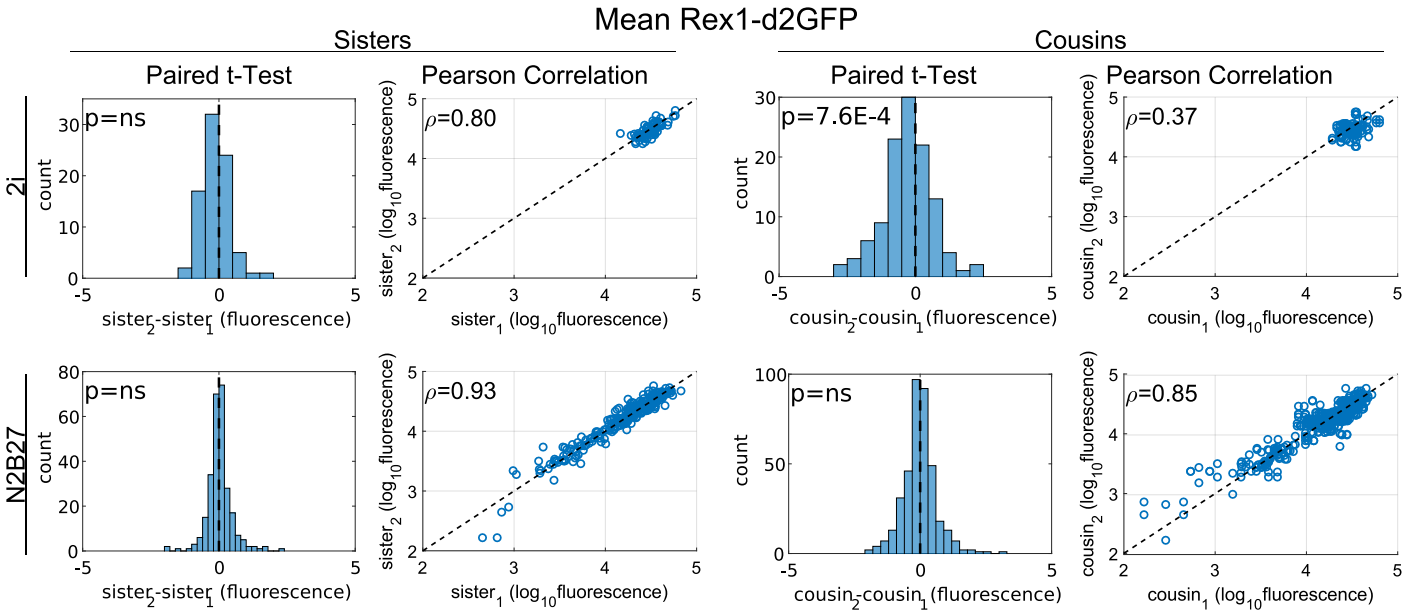


Fig. 5.26 Genealogical analysis of mean lifetime Rex1-GFPd2 signal. Comparison between sisters and cousins, for both 2i ES cells and N2B27 cells. The paired t-Test shows the distribution of the difference between two cells with the associated p-value (p). The Pearson correlation shows a log-log scatter plot of the paired cells with a reference line (black dashed line, slope = 1, intercept = 0) and the correlation coefficient (ρ). ns = not significant. Biological Replicate 1 is shown here, with all data summarized in Table 5.8

Table 5.8 Summary of mean Rex1-GFPd2 analysis for 2i ES cells and N2B27 cells. (t-test significance with p-value, Pearson correlation with p-value, number of cells, and number of clones.)

Genealogical Analysis of Mean Rex1-GFPd2												
	2i						**N2B27**					
	t-Test	p-value	ρ	p-value	N_{cell}	N_{clone}	t-Test	p-value	ρ	p-value	N_{cell}	N_{clone}
ss	ns	9.8E-2	0.80	<1.0E-4	82	17	ns	0.29	0.94	<1.0E-4	271	62
	ns	1.0	0.86	<1.0E-4	25	10	ns	0.87	0.95	<1.0E-4	101	30
cc	***	7.6E-4	0.38	<1.0E-4	115	16	ns	0.75	0.86	<1.0E-4	386	45
	ns	2.8E-2	0.71	<1.0E-4	28	5	ns	0.65	0.89	<1.0E-4	136	24
md	***	4.2E-4	0.62	<1.0E-4	144	17	****	<1E-4	0.72	<1.0E-4	478	52
	ns	1.7E-2	0.72	<1.0E-4	31	6	****	<1E-4	0.78	<1.0E-4	171	26
gd	****	<1E-4	0.34	1.8E-3	83	13	****	<1E-4	0.00	ns	298	33
	ns	3.2E-3	0.79	ns	10	2	****	<1E-4	0.43	<1.0E-4	80	15

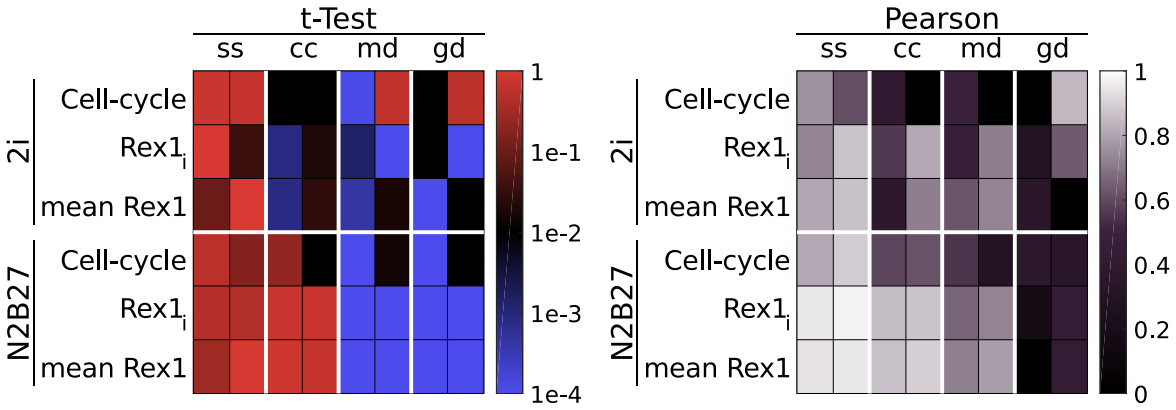


Fig. 5.27 Heat-maps of t-Test and Pearson correlation for genealogical relations of cell-cycle duration (Cell-cycle), initial Rex1-GFPd2 signal (Rex1_i), and mean life time Rex1-GFPd2 signal (mean Rex1). The columns show biological replicates. The colorbar of t-Test shows the p-value where shades of blue are significant and shades of red are not significant. The colorbar for Pearson's correlation is the correlation coefficient ρ. (ss = sisters, cc = first-cousins, md = mother/daughter, and gd = grandmother/granddaughter).

5.4.4 Summary

The global distribution of Rex1-GFPd2 expression levels shows the many hours over which this reporter is down-regulated. This behavior indicates the asynchronicity of the exit from the naïve state. This is further shown by transforming these data into the fraction of cells that have exited at time t. These data were compared against the bulk-culture assays to demonstrate that the culture from the long-term single-cell imaging assays are healthy and that the process of exit is occurring asynchronously.

The mechanism of how Rex1-GFPd2 is regulated has been investigated by shifting clones in time. A sudden and irreversible inactivation of the reporter is seen which is indicative of a genetic switch mechanism. When the kinetics of exit are transformed into the fraction of exited cells, a tight hockey stick curve is seen. This demonstrates a high level of intraclonal synchronicity.

Finally, it was shown that as ES cells exit naïve pluripotency in a highly coordinated manner. The correlation between cells within a generation is extremely high and highly significant, and more correlated than the corresponding relations for ES cells self-renewing. Cells are born with the same amount of Rex1-GFPd2, seen in the initial Rex1-GFPd2 signal, and they maintain a very similar level over their lives, seen in the mean Rex1-GFPd2 signal. Thus leading to the conclusion that ES cells predominantly use symmetric divisions during the exit from naïve pluripotency. It does not exclude the possibility of asymmetric divisions, but it does imply that if asymmetric divisions do occur, they occur at low frequencies.

5.5 Discussion

Withdrawal of 2i from ES cells results in the release from naïve pluripotency. This allows ES cells to begin the process of differentiation by global changes in the transcriptome and the methylome (Kalkan et al., 2017). Withdrawal of these two inhibitors also comes with directly observable behavioral changes. The technique of long-term single-cell imaging and tracking allowed for thorough examination of the morphology, motility, cell-cycle, and the kinetics of Rex1-GFPd2 down-regulation.

A general trend in morphology is that cells become smaller as the colony becomes larger, regardless of culture condition. However, the sizes of 2i ES cells are consistently smaller than that of N2B27 cells. The rate of cell size decrease slows with the onset of exit. A further distinction arises in cell motility. 2i ES cells remain at a constant low motility, whereas N2B27 cells become more motile the further they move away from naïve pluripotency. Taken together, the divergence of size and motility results in morphologically distinct colonies. 2i ES cell colonies are small and compact while N2B27 colonies are large and diffuse.

Quantification of individual cells shows an acceleration in the cell-cycle duration of N2B27 cells. These data were then used to match the growth curves of the single-cell data to the bulk-culture data in Ch. 3. This confirms the relative health of the single-cell imaging culture. Examination of the time in the cell-cycle at which self-renewal signals were withdrawn shows no clear effect on kinetics of Rex1-GFPd2 down-regulation. Analysis of the cell-cycle concluded with the examination of the heritability of cell-cycle duration. It is observed that a high level of correlation exist within a generation (i.e. sisters) but not across generations (i.e. mothers and daughters.). It was also determined that the cell-cycle duration of N2B27 cell are more strongly correlated than that of 2i ES cells.

The kinetics of Rex-GFPd2 down regulation were observed to be asynchronous in real time. However, once lineages where shifted in time, an exponential decay trend appeared. This is consistent with a sudden and irreversible inactivation of the reporter, i.e. a switch mechanism. Genealogical analysis of the initial and mean Rex1-GFPd2 expression levels showed high levels of correlation for sisters in 2i and N2B27, and for first-cousins only in N2B27. These data suggest that ES cells predominantly use symmetric cell divisions during the exit from naïve pluripotency.

These data present a model for the down regulation of the pluripotency marker Rex1-d2GFP. First, the process of exit is initiated in an asynchronous fashion, independent of the cell cycle. When cells exit the naïve state, Rex1-d2GFP is shut off as a genetic switch. The level of intraclonal synchronicity and correlation is greater than 2i ES cells, indicating

that ES cells divide symmetrically as they exit. Therefore, the picture that emerges for exit from naïve pluripotency is that upon the withdrawal of self-renewal signals, ES cells exit asynchronously via symmetric divisions.

Chapter 6

General Discussion

"... I HAVE ADDED SOMETHING TO KNOWLEDGE, AND HELPED OTHERS TO ADD MORE; AND THAT THESE SOMETHINGS HAVE A VALUE WHICH DIFFERS IN DEGREE ONLY, AND NOT IN KIND, FROM THAT OF THE CREATIONS OF THE GREAT MATHEMATICIANS, OR OF ANY OF THE OTHER ARTISTS, GREAT OR SMALL, WHO HAVE LEFT SOME KIND OF MEMORIAL BEHIND THEM."

- G.H. HARDY, 1940

6.1 Findings

I have set out to examine the dynamics and factors affecting the exit from naïve pluripotency. Distinct changes in motility, morphology, and cell-cycle duration were observed in ES cells as they exit naïve pluripotency. Genealogical analyzes have shown evidence that cell division predominantly occurs symmetrically during this process. When Rex1-GFPd2 is down-regulated, it is irreversible and it occurs as a single step switch. And finally, that asynchronicity exists on a population level, while high levels of synchronicity exist within a single clone. These findings are summarized in Fig. 6.1.

6.1.1 How does ES cell behavior change during exit?

Single-cell snapshots were used to determine the doubling time for ES cells and cells exiting pluripotency. In the chemically defined 2i system, no difference was observed in the doubling time between 2i ES cells and N2B27 cells. However, the serum based cultures showed that the doubling time of serum cells was shorter than LPS ES cells. But, doubling time is a compound metric which cannot parse proliferation and cell death, therefore these data

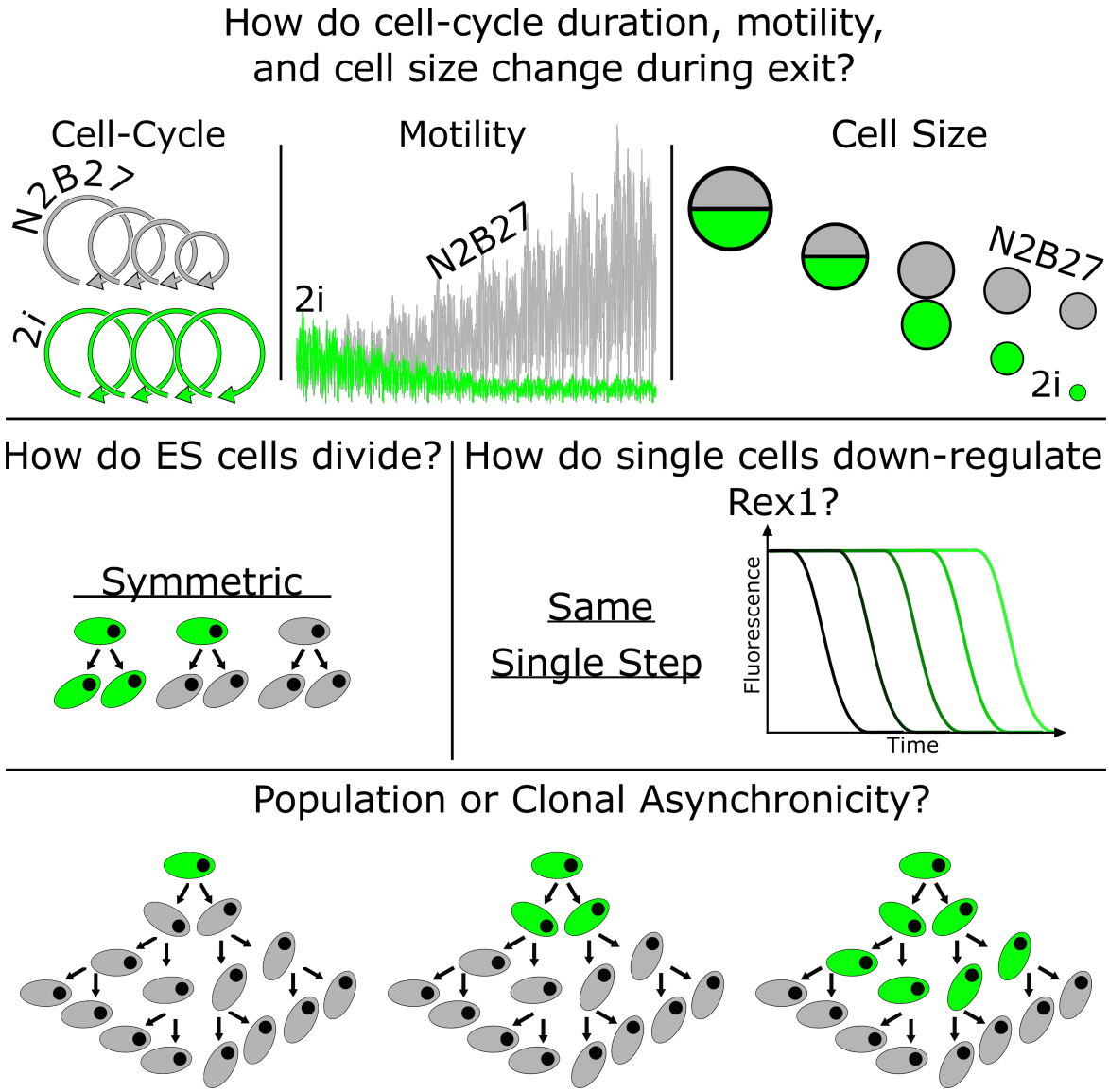


Fig. 6.1 (A) Generational analysis has shown that as ES cells self-renew they maintain a constant cell-cycle duration, exhibit decreasing levels of motility, and occyoy less area over subsequent generations. As ES cells exit self-renewal, their cell-cycle becomes ever shorter, they exhibit increasing motility, and while the area they occupy becomes smaller, they remain consistently larger than ES cells. (B) Genealogical comparisons have shown high levels of correlation among exiting ES cells, providing strong evidence that divisions are predominantly symmetric. (C) Time-shifted genetic trajectory multiple alignment of lineages revealed a sudden and irreversible inactivation of the Rex1-GFPd2 reporter indicating that down-regulation proceeds as a single step switch. (D) B and C taken together show a high level of intraclonal synchronicity, meaning that asynchronicity arises from individual clones exiting at different times.

cannot determine whether the faster doubling time in serum was unique to that system, or if increased rates of cell-cycling is a universal feature of exit, where apoptosis was masking this feature in N2B27 cells. Thus, long-term single-cell imagining and tracking was used to directly observe the cell-cycle duration of 2i ES cells and N2B27 cells.

The cell-cycle becomes increasingly shorter as cells move further away from naïve pluripotency (Fig. 6.1 A, left). Analysis of long-term single-cell tracking showed that the cell-cycle of N2B27 cell was approximately one hour shorter than 2i ES cells. Further generational analysis showed a progressive shortening of cell-cycle duration for N2B27 cell. This was up to three hours shorter than 2i ES cells, which maintained their cell-cycle duration over generations. Thus leading to the conclusion that the rate of cell-cycling increases during the exit from naïve pluripotency. Even though the bulk culture proliferation assays for 2i showed no significant difference in doubling time between 2i ES cells and N2B27 cells, this could be attributed to increased apoptosis during exit. The shortening of the cell-cycle for exiting cells is most likely connected to the withdrawal of Mek inhibitor PD03 and the reactivation of the Fgf signaling pathway. Disruption of Fgf signaling results in disruption of proliferation, therefore the return of normal Fgf signaling allows for normal proliferation to occur (Lanner and Rossant, 2010). *In vitro*, ter Huurne et al., 2017, attribute a longer G1 phase in 2i ES cells, as compared to serum/Lif cells, to the presence of PD03. Thus, it stands to reason that the shortening of the cell-cycle of N2B27 cells is a result of shortening of the G1 phase in the absence of PD03. However, that same study also reports a shorter G2 phase in 2i ES cells than in serum/Lif, so a full explanation of the cell-cycle acceleration may be non-trivial. Finally, increased rates of proliferation are also seen in the embryo during gastrulation. The ICM develops zones of proliferation where the cell-cycle becomes as short as four hours (Snow, 1977).

It would be interesting to further explore what phases in the cell cycle change, and the potential effect these changes would have on lineage bias. In hPSCs, Smad signaling differentially regulates lineage specification throughout the cell cycle (Pauklin and Vallier 2013, 2014). In the early G1 phase, Smad2/3 locates to the nucleus where it binds and activates endoderm genes. In late G1 and G1/S phase Smad2/3 is excluded from the nucleus via phosphorylation by CDK3/4. At this point in the cell-cycle, endoderm specification is prohibited, providing a permissive state for neuroectoderm induction (Pauklin and Vallier 2013, 2014). The shortened G1, phase upon withdrawal of PD03 observed by ter Huurne et al., 2017, could provide a means to skip past a permissive state for endoderm specification to one that allows neural induction. This could be explored by observing directed differentiation with long-term single-cell tracking of a FUCCI system (Sakaue-Sawano et al., 2008a,b),

paired with an endpoint immunostaining for lineage specific markers. The point of this would be to determine if alterations in length of cell-cycle phases are permissive or prohibitive for specific lineages. In the first instance, differentiation regimes towards neural lineages and endodermal lineages should be carried out to determine if the phases of the cell-cycle change on the basis of exit conditions. A lengthening of a permissive phase may allow cells to differentiate towards a certain lineage with greater efficiency. While, on the other hand, a shortening of a permissive phase would be a way to bypass an alternative fate.

A connection between morphology and motility is also observed in exiting ES cells. A distinct morphological difference exists between ES cell colonies and differentiated cell colonies. The observation that ES cell colonies are compact and differentiated colonies are diffuse has been reported here and elsewhere (Andersson-Rolf et al., 2017; Wray et al., 2011). The withdrawal of Gsk3 inhibitor Chiron results in depletion of β-catenin, a stabilizer of adherens junctions, which is crucial for cell-cell contact (Tian et al., 2011). RNA-seq shows a concomitant down-regulation of other key adhesion molecules Jam2, Pecam1, and Icam1 during the transition to exit (Kalkan et al., 2017); indicating that a loss of cell-cell contact could be expected. Generational analysis of motility show that 2i ES cells become less motile while N2B27 cells become more active (Fig. 6.1 A, middle). Increased motility could be linked to the reactivation of Erk signaling which promotes activation of focal adhesions and, more generally, locomotion (Tanimura and Takeda, 2017).

Also, the area of both 2i ES cells and N2B27 cells becomes progressively smaller, however, 2i ES cells are consistently smaller than their exiting counterparts (Fig. 6.1 A, right). This suggests that the ability of ES cells to maintain cell-cell contact keeps them confined to their neighborhood. I would further suggest that at each round of cell division, ES cells retain less area, so they should tend to grow vertically. This is the opposite of what is seen during exit. Decreased cell-cell contact results in more motile cells that are unable to hold onto their neighbors, so they lay flat in the dish. Once again, during gastrulation, migration is observed during primitive streak formation. These data suggest that some form of mesenchymal migration is occurring *in vitro* (Scarpa and Mayor, 2016).

The withdrawal of the two inhibitors leads to the reactivation of Erk activity and the suppression β-catenin. The result is that major changes occur in the transcriptome, methylome, and metabolism of ES cells as they move towards the exit from naïve pluripotency (Kalkan et al., 2017). Long-term single-cell imaging and tracking have revealed that these changes result in three behaviors that deviate from self-renewing naïve ES cells: cells with greater areas, increased motility, and shorter cell-cycles.

6.1.2 How do ES cells divide during exit: symmetric, asymmetric, or mixed?

It was classically thought that adult stem cells used asymmetric divisions to fine-tune the balance between self-renewal and differentiation (Simons and Clevers, 2011). Yet some systems also employ a strategy of asynchronous symmetric divisions (Simons and Clevers, 2011) to maintain stem cell populations in adult tissues over the lifetime of the organism. Habib et al., 2013, have shown that it is indeed possible to induce asymmetric division *in vitro* by the introduction of extrinsic signals, but are there instances of non-forced asymmetric division?

Here I have quantified division symmetry by conducting genealogical comparisons for cell-cycle duration, initial Rex1-GFPd2 signal, and mean Rex1-GFPd2 signal. High levels of correlation were seen for sisters in 2i and N2B27, and also for first-cousins only in N2B27 conditions. Intergenerational comparisons between generations showed low to no correlation. In all cases the level of correlation was higher in N2B27 cells than 2i ES cells. This suggests that potency is inherited symmetrically during exit from naïve pluripotency (Fig. 6.1 B).

It has recently been suggested that during differentiation of human ES cells, division occurs asymmetrically in 30-50% of cases (Brown et al., 2017). However, human ES cells are more developmentally advanced, and primed for differentiation, as compared to mouse naïve ES cells. In this study, Brown et al., 2017, used single-cell imaging to quantify the expression levels of pluripotency markers for cells after the first division. They define an asymmetric division as one in which the expression level is one cell is 1.5 times greater than the other. In addition to the single generation analysis and the arbitrary threshold, they did not report the number cells tracked nor did they employ any statical methods to strengthen their claim. I believe this case highlights a need for rigorous analytical methods to interrogate these data.

6.1.3 How do ES cells down-regulate Rex1?

The serum/Lif culture system displays heterogeneous expression of key pluripotency markers, such as the bimodal expression of Nanog (Chambers et al., 2007; Chickarmane et al., 2006; Toyooka et al., 2008). This heterogeneity is indeed a dynamical system where negative feedback induces stochastic switching between Nanog high and low states via a hysteresis loop (Herberg et al., 2016). In the dynamic heterogeneity model, it is thought that ES cells must first enter a low state, where they have the ability to explore the space of lineage choices, which primes them for eventual specification (Moris et al., 2016). On the other hand, ES

cells in 2i display a more uniform expression of the naïve factors (Kalkan et al., 2017; Wray et al., 2011; Ying et al., 2008). Exit does proceed in an asynchronous fashion but the 2i system does not exhibit any state switching during self-renewal. Exit in this scenario has been hypothesized to proceed by a progression of consecutive phases, where ES cells are not prepared to execute lineage specification and must first proceed through a formative maturation phase prior to priming and eventual specification (Smith, 2017).

In this work, no instances were observed where a differentiating cell spontaneously reactivated Rex1-GFPd2. Given the unidirectionality of development, it does not seem likely that there should be spontaneous reacquisition of the pluripotent identity. Rather, the onset of exit is associated with a sudden and irreversible inactivation of the Rex1-GFPd2 reporter (Fig. 6.1 C). Time-shifted genetic trajectory multiple alignment of lineages further revealed a global trend of normal expression followed by the sudden onset of exponential decay, which is indicative of a genetic switch. This sudden and irreversible inactivation of the Rex1-GFPd2 reporter during exit fits the phased progression model of ES cell differentiation suggested by Smith, 2017, and the unidirectional nature of development.

6.1.4 Does the asynchronicity of exit arise on the population or the clonal level?

The asynchronicity of exit from naïve pluripotency has previously been noted (Betschinger et al., 2013; Kalkan et al., 2017; Leeb et al., 2014; Lowell et al., 2006; Zhou et al., 2013). Here I have established that exit from naïve pluripotency occurs asynchronously on the population level in both 2i and Lif & PD03 in serum. However, the origin of this asynchronicity was still not clear from the bulk-culture data. Possibilities included reversible switching such as that seen in serum/Lif or the possibility of lineage branch asymmetries. The long-term single-cell imaging data suggests that neither of these can be the case. Rather, when a single lineage tree, generated from an individual single cell (i.e. a clonal population), is analyzed, synchronous behaviors are observed. This indicated that variability between clones cause the macroscopic asynchronous behavior observed at the population level (Fig. 6.1 D).

It is unclear from these data why this is the case. The cells were not synchronized in any manner prior to self-renewal signal withdrawal, but the possibility of cell-cycle timing was eliminated by performing correlation analysis on the "time of first division" and "the time of first exit". This showed that the time in the cell-cycle at which point self-renewal signals are withdrawn had no effect on the kinetics of Rex1-GFPd2 down-regulation. Furthermore, exit was observed to occur within different generations, meaning there is no preset requirement

on the number of divisions a cell must undergo prior to exiting naïve pluripotency. Zhou et al., 2013, have suggested the role of Hes1 in the desynchronization of differentiation for the serum/Lif culture system by modulation of Stat3 signaling (Kobayashi and Kageyama, 2010; Zhou et al., 2013). An immunofluorescence time-course, costaining for Rex1 and Hes1 in ES cells, during the exit from the naïve state would be a good start to determine whether Hes1 is a party to the observed desynchronization. If it is the case that Hes1 is in some way involved, upregulation in Hes1 should be observed with the down regulation of Rex1.

6.2 Futures

6.2.1 The Synchronicity of *In Vivo* Exit From Naïve Pluripotency

Segregation of the E3.5 ICM into epiblast and primitive endoderm occurs by incremental asynchronous allocation of cells from a common progenitor pool (Saiz et al., 2016). This is thought to provide a basis for the regulation of tissue size. Asynchronous behavior is also seen in the down-regulation of Nanog peri-implantation between the E4.5 and E4.7 embryo (Acampora et al., 2016). However, these data come from snap-shots of embryos and it is not clear how this process proceeds in real time.

The population asynchronicity seen *in vitro* may be an indicator for the behavior of the epiblast at this stage of development, given the genetic similarity between the two cell types (Boroviak et al., 2014; Cockburn and Rossant, 2010) and now the similarity in the time scale of dissolution of the naïve network in the embryo and in ES cells. Naïve pluripotency emerges in the final day prior to its dissolution at implantation (Boroviak and Nichols, 2014) and exit in the 2i system, while slightly longer, has been measured to take place over 31.4 ± 13.6 hours. It would be interesting to further investigate if incremental allocation of cell fates is a feature that persists throughout development as a mechanism for tissue size regulation.

However, not only is this stage of development difficult to capture, but it is also difficult to live-image (personal communication, Nichols lab). Three dimensional long-term single-cell imaging needs to be coupled with a tracking platform that can segment volumes. While technologies like light-sheet microscopy are capable of producing sufficient images, the density of the epiblast makes individual cell identification difficult.

6.2.2 Neighborhood Networks

The data sets presented here contain additional information that has not been investigated in this thesis. One aspect in particular that has not been explored is cellular neighborhoods. A

network of neighbors can be constructed using the segmented regions of interest to determine which cells are in contact. These networks can be used to look at how the time evolution of different attributes are affected by cell-cell contact. Such analyzes could be complemented by the in *vivo* studies mentioned in the previous section. Of note, there might be a significant difference in the number of cells neighboring pluripotent cells in the embryo compared to the *in vitro* system.

6.2.3 A Citizen Science Solution to the Single-Cell Tracking Bottleneck

Accurate segmentation and lineage tracing is still a bottleneck in processing long-term single-cell imaging data. The ultimate tool would be fully automated software which can operate with minimal human supervision. This aspiration will hopefully become a reality with improvements in imaging technology, tracking algorithms, and machine vision. Until then, I propose a solution in the form of citizen science. Much like the NASA "Clickworkers" mapping the craters of Mars, I think that a community of people tracking cell lineages is a reasonable stop-gap.

The performance of single-cell tracking does not require a high level of technical proficiency. For each image, all cells must be identified, segmented, and prescribed a fate (divide, die, unknown). Rather, the difficulty comes with the months it takes an individual to analyze thousands of frames of footage. This community tracking project should be accessible through an internet gateway. The "tracker" would complete a training module which would be comprised of a series of pre-tracked lineages that become progressively complex. Once that person completes the training module, they can start tracking randomly selected lineages. Many people will then track the same lineages from which a consensus lineage can be constructed. This would require managing the large data sets from long-term single-cell imaging, engineering a tracking and analysis software that can be hosted online, and people. An ideal community would be sixth form/high school students. In addition to being acknowledged on any publications, students would receive instruction in biology, computer science, mathematics, and programming.

References

Acampora, D., Omodei, D., Petrosino, G., Garofalo, A., Savarese, M., Nigro, V., Di Giovannantonio, L. G., Mercadante, V., and Simeone, A. (2016). Loss of the otx2-binding site in the nanog promoter affects the integrity of embryonic stem cell subtypes and specification of inner cell mass-derived epiblast. *Cell reports*, 15(12):2651–2664.

Ambrosetti, D.-C., Schöler, H. R., Dailey, L., and Basilico, C. (2000). Modulation of the activity of multiple transcriptional activation domains by the dna binding domains mediates the synergistic action of sox2 and oct-3 on the fibroblast growth factor-4enhancer. *Journal of Biological Chemistry*, 275(30):23387–23397.

Andersson-Rolf, A., Mustata, R. C., Merenda, A., Kim, J., Perera, S., Grego, T., Andrews, K., Tremble, K., Silva, J. C., Fink, J., et al. (2017). One-step generation of conditional and reversible gene knockouts. *Nature methods*, 14(3):287–289.

Arman, E., Haffner-Krausz, R., Chen, Y., Heath, J. K., and Lonai, P. (1998). Targeted disruption of fibroblast growth factor (fgf) receptor 2 suggests a role for fgf signaling in pregastrulation mammalian development. *Proceedings of the National Academy of Sciences*, 95(9):5082–5087.

Avilion, A. A., Nicolis, S. K., Pevny, L. H., Perez, L., Vivian, N., and Lovell-Badge, R. (2003). Multipotent cell lineages in early mouse development depend on sox2 function. *Genes & development*, 17(1):126–140.

Beddington, R. and Robertson, E. (1989). An assessment of the developmental potential of embryonic stem cells in the midgestation mouse embryo. *Development*, 105(4):733–737.

Berman, H., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T., Weissig, H., Shindyalov, I., and Bourne, P. (2000). The protein data bank. *Nucleic Acids Research*, pages 235–242.

Betschinger, J. (2017). Charting developmental dissolution of pluripotency. *Journal of molecular biology*, 429(10):1441–1458.

Betschinger, J. and Knoblich, J. A. (2004). Dare to be different: asymmetric cell division in drosophila, c. elegans and vertebrates. *Current biology*, 14(16):R674–R685.

Betschinger, J., Nichols, J., Dietmann, S., Corrin, P. D., Paddison, P. J., and Smith, A. (2013). Exit from pluripotency is gated by intracellular redistribution of the bhlh transcription factor tfe3. *Cell*, 153(2):335–347.

Blanchard, G. B., Kabla, A. J., Schultz, N. L., Butler, L. C., Sanson, B., Gorfinkiel, N., Mahadevan, L., and Adams, R. J. (2009). Tissue tectonics: morphogenetic strain rates, cell shape change and intercalation. *Nature methods*, 6(6):458–464.

Boroviak, T., Loos, R., Bertone, P., Smith, A., and Nichols, J. (2014). The ability of inner-cell-mass cells to self-renew as embryonic stem cells is acquired following epiblast specification. *Nature cell biology*, 16(6):513–525.

Boroviak, T., Loos, R., Lombard, P., Okahara, J., Behr, R., Sasaki, E., Nichols, J., Smith, A., and Bertone, P. (2015). Lineage-specific profiling delineates the emergence and progression of naive pluripotency in mammalian embryogenesis. *Developmental cell*, 35(3):366–382.

Boroviak, T. and Nichols, J. (2014). The birth of embryonic pluripotency. *Phil. Trans. R. Soc. B*, 369(1657):20130541.

Boyer, L. A., Lee, T. I., Cole, M. F., Johnstone, S. E., Levine, S. S., Zucker, J. P., Guenther, M. G., Kumar, R. M., Murray, H. L., Jenner, R. G., et al. (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. *cell*, 122(6):947–956.

Bradley, A., Evans, M., Kaufman, M. H., and Robertson, E. (1984). Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature*, 309(5965):255–256.

Brinster, R. L. (1974). The effect of cells transferred into the mouse blastocyst on subsequent development. *Journal of Experimental Medicine*, 140(4):1049–1056.

Brown, K., Loh, K. M., and Nusse, R. (2017). Live imaging reveals that the first division of differentiating human embryonic stem cells often yields asymmetric fates. *Cell Reports*, 21(2):301–307.

Brown, W. K. and Wohletz, K. H. (1995). Derivation of the weibull distribution based on physical principles and its connection to the rosin–rammler and lognormal distributions. *Journal of Applied Physics*, 78(4):2758–2763.

Buecker, C., Srinivasan, R., Wu, Z., Calo, E., Acampora, D., Faial, T., Simeone, A., Tan, M., Swigut, T., and Wysocka, J. (2014). Reorganization of enhancer patterns in transition from naive to primed pluripotency. *Cell stem cell*, 14(6):838–853.

Burdon, T., Chambers, I., Stracey, C., Niwa, H., and Smith, A. (1999a). Signaling mechanisms regulating self-renewal and differentiation of pluripotent embryonic stem cells. *Cells Tissues Organs*, 165(3-4):131–143.

Burdon, T., Stracey, C., Chambers, I., Nichols, J., and Smith, A. (1999b). Suppression of shp-2 and erk signalling promotes self-renewal of mouse embryonic stem cells. *Developmental biology*, 210(1):30–43.

Cannon, D., Corrigan, A. M., Miermont, A., McDonel, P., and Chubb, J. R. (2015). Multiple cell and population-level interactions with mouse embryonic stem cell heterogeneity. *Development*, 142(16):2840–2849.

Carter, A. C., Davis-Dusenbery, B. N., Koszka, K., Ichida, J. K., and Eggan, K. (2014). Nanog-independent reprogramming to ipscs with canonical factors. *Stem cell reports*, 2(2):119–126.

Cartwright, P., McLean, C., Sheppard, A., Rivett, D., Jones, K., and Dalton, S. (2005). Lif/stat3 controls es cell self-renewal and pluripotency by a myc-dependent mechanism. *Development*, 132(5):885–896.

Chambard, J.-C., Lefloch, R., Pouysségur, J., and Lenormand, P. (2007). Erk implication in cell cycle regulation. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, 1773(8):1299–1310.

Chambers, I. (2004). The molecular basis of pluripotency in mouse embryonic stem cells. *Cloning and stem cells*, 6(4):386–391.

Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S., and Smith, A. (2003). Functional expression cloning of nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell*, 113(5):643–655.

Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., Vrana, J., Jones, K., Grotewold, L., and Smith, A. (2007). Nanog safeguards pluripotency and mediates germline development. *Nature*, 450(7173):1230–1234.

Chambers, I. and Smith, A. (2004). Self-renewal of teratocarcinoma and embryonic stem cells. *Oncogene*, 23(43):7150–7160.

Chambers, I. and Tomlinson, S. R. (2009). The transcriptional foundation of pluripotency. *Development*, 136(14):2311–2322.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V. B., Wong, E., Orlov, Y. L., Zhang, W., Jiang, J., et al. (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell*, 133(6):1106–1117.

Chenn, A. and McConnell, S. K. (1995). Cleavage orientation and the asymmetric inheritance of notchl immunoreactivity in mammalian neurogenesis. *Cell*, 82(4):631–641.

Chew, J.-L., Loh, Y.-H., Zhang, W., Chen, X., Tam, W.-L., Yeap, L.-S., Li, P., Ang, Y.-S., Lim, B., Robson, P., et al. (2005). Reciprocal transcriptional regulation of pou5f1 and sox2 via the oct4/sox2 complex in embryonic stem cells. *Molecular and cellular biology*, 25(14):6031–6046.

Chickarmane, V., Troein, C., Nuber, U. A., Sauro, H. M., and Peterson, C. (2006). Transcriptional dynamics of the embryonic stem cell switch. *PLoS computational biology*, 2(9):e123.

Clevers, H. (2005). Stem cells, asymmetric division and cancer. *Nature genetics*, 37(10):1027.

Cockburn, K. and Rossant, J. (2010). Making the blastocyst: lessons from the mouse. *The Journal of clinical investigation*, 120(4):995.

Conboy, I. M. and Rando, T. A. (2002). The regulation of notch signaling controls satellite cell activation and cell fate determination in postnatal myogenesis. *Developmental cell*, 3(3):397–409.

Coutu, D. L. and Schroeder, T. (2013). Probing cellular processes by long-term live imaging–historic problems and current solutions. *J Cell Sci*, 126(17):3805–3815.

Davies, O. R., Lin, C.-Y., Radzisheuskaya, A., Zhou, X., Taube, J., Blin, G., Waterhouse, A., Smith, A. J., and Lowell, S. (2013). Tcf15 primes pluripotent cells for differentiation. *Cell reports*, 3(2):472–484.

Dhaliwal, N. K., Miri, K., Davidson, S., El Jarkass, H. T., and Mitchell, J. A. (2018). Klf4 nuclear export requires erk activation and initiates exit from naive pluripotency. *Stem cell reports*.

Diwan, S. B. and Stevens, L. C. (1976). Development of teratomas from the ectoderm of mouse egg cylinders. *Journal of the National Cancer Institute*, 57(4):937–942.

Doble, B. W., Patel, S., Wood, G. A., Kockeritz, L. K., and Woodgett, J. R. (2007). Functional redundancy of gsk-3α and gsk-3β in wnt/β-catenin signaling shown by using an allelic series of embryonic stem cell lines. *Developmental cell*, 12(6):957–971.

Doe, C. Q. and Bowerman, B. (2001). Asymmetric cell division: fly neuroblast meets worm zygote. *Current opinion in cell biology*, 13(1):68–75.

Dunn, S.-J., Martello, G., Yordanov, B., Emmott, S., and Smith, A. (2014). Defining an essential transcription factor program for naive pluripotency. *Science*, 344(6188):1156–1160.

Eliot, T. S. (1943). *Four quartets*. Harcourt.

Etzrodt, M., Endele, M., and Schroeder, T. (2014). Quantitative single-cell approaches to stem cell research. *Cell stem cell*, 15(5):546–558.

Evans, M. J. and Kaufman, M. H. (1981). Establishment in culture of pluripotential cells from mouse embryos. *Nature*, 292(5819):154–156.

Filipczyk, A., Marr, C., Hastreiter, S., Feigelman, J., Schwarzfischer, M., Hoppe, P. S., Loeffler, D., Kokkaliaris, K. D., Endele, M., Schauberger, B., et al. (2015). Network plasticity of pluripotency transcription factors in embryonic stem cells. *Nature cell biology*, 17(10):1235.

Fréchet, M. R. (1927). Sur la loi de probabilité de l'écart maximum. *Annales de la Société Polonaise de Mathematique*, 6:93–116.

Glauche, I., Herberg, M., and Roeder, I. (2010a). Nanog variability and pluripotency regulation of embryonic stem cells - insights from a mathematical model analysis. *PLOS ONE*, 5(6):1–12.

Glauche, I., Herberg, M., and Roeder, I. (2010b). Nanog variability and pluripotency regulation of embryonic stem cells-insights from a mathematical model analysis. *PLoS One*, 5(6):e11238.

Goedhart, J., Von Stetten, D., Noirclerc-Savoye, M., Lelimousin, M., Joosen, L., Hink, M. A., Van Weeren, L., Gadella Jr, T. W., and Royant, A. (2012). Structure-guided evolution of cyan fluorescent proteins towards a quantum yield of 93%. *Nature communications*, 3:751.

Gönczy, P. and Rose, L. S. (2005). Asymmetric cell division and axis formation in the embryo. *WormBook*, 15:1–20.

Guo, G., Huang, Y., Humphreys, P., Wang, X., and Smith, A. (2011). A piggybac-based recessive screening method to identify pluripotency regulators. *PLoS One*, 6(4):e18189.

Guo, G., Huss, M., Tong, G. Q., Wang, C., Sun, L. L., Clarke, N. D., and Robson, P. (2010). Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. *Developmental cell*, 18(4):675–685.

Habib, S. J., Chen, B.-C., Tsai, F.-C., Anastassiadis, K., Meyer, T., Betzig, E., and Nusse, R. (2013). A localized wnt signal orients asymmetric stem cell division in vitro. *Science*, 339(6126):1445–1448.

Hamazaki, T., Kehoe, S. M., Nakano, T., and Terada, N. (2006). The grb2/mek pathway represses nanog in murine embryonic stem cells. *Molecular and cellular biology*, 26(20):7539–7549.

Hardy, G. H. (1940). *A Mathematician's Apology*. Cambridge University Press.

Harland, R. (2000). Neural induction. *Current opinion in genetics & development*, 10(4):357–362.

Hayashi, K., Lopes, S. M., Tang, F., and Surani, M. A. (2008). Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states. *Cell stem cell*, 3(4):391–401.

Herberg, M., Glauche, I., Zerjatke, T., Winzi, M., Buchholz, F., and Roeder, I. (2016). Dissecting mechanisms of mouse embryonic stem cells heterogeneity through a model-based analysis of transcription factor dynamics. *Journal of The Royal Society Interface*, 13(117).

Herberg, M., Kalkan, T., Glauche, I., Smith, A., and Roeder, I. (2014). A model-based analysis of culture-dependent phenotypes of mescs. *PLOS ONE*, 9(3):1–12.

Herberg, M. and Roeder, I. (2015). Computational modelling of embryonic stem-cell fate control. *Development*, 142(13):2250–2260.

Hilsenbeck, O., Schwarzfischer, M., Skylaki, S., Schauberger, B., Hoppe, P. S., Loeffler, D., Kokkaliaris, K. D., Hastreiter, S., Skylaki, E., Filipczyk, A., et al. (2016). Software tools for single-cell tracking and quantification of cellular and molecular properties. *Nature biotechnology*, 34(7):703–706.

Jaenisch, R. and Young, R. (2008). Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. *Cell*, 132(4):567–582.

Kaji, K., Caballero, I. M., MacLeod, R., Nichols, J., Wilson, V. A., and Hendrich, B. (2006). The nurd component mbd3 is required for pluripotency of embryonic stem cells. *Nature cell biology*, 8(3):285–292.

Kaji, K., Nichols, J., and Hendrich, B. (2007). Mbd3, a component of the nurd co-repressor complex, is required for development of pluripotent cells. *Development*, 134(6):1123–1132.

Kalkan, T., Olova, N., Roode, M., Mulas, C., Lee, H. J., Nett, I., Marks, H., Walker, R., Stunnenberg, H. G., Lilley, K. S., et al. (2017). Tracking the embryonic stem cell transition from ground state pluripotency. *Development*, pages dev–142711.

Kalkan, T. and Smith, A. (2014). Mapping the route from naive pluripotency to lineage specification. *Phil. Trans. R. Soc. B*, 369(1657):20130540.

Kalmar, T., Lim, C., Hayward, P., Muñoz-Descalzo, S., Nichols, J., Garcia-Ojalvo, J., and Arias, A. M. (2009). Regulated fluctuations in nanog expression mediate cell fate decisions in embryonic stem cells. *PLoS biology*, 7(7):e1000149.

Kim, J., Chu, J., Shen, X., Wang, J., and Orkin, S. H. (2008). An extended transcriptional network for pluripotency of embryonic stem cells. *Cell*, 132(6):1049–1061.

Kim, J. D., Kim, H., Ekram, M. B., Yu, S., Faulk, C., and Kim, J. (2011). Rex1/zfp42 as an epigenetic regulator for genomic imprinting. *Human molecular genetics*, 20(7):1353–1362.

Kleinsmith, L. J. and Pierce, G. B. (1964). Multipotentiality of single embryonal carcinoma cells. *Cancer research*, 24(9):1544–1551.

Kobayashi, T. and Kageyama, R. (2010). Hes1 regulates embryonic stem cell differentiation by suppressing notch signaling. *Genes to Cells*, 15(7):689–698.

Kolodziejczyk, A. A., Kim, J. K., Tsang, J. C., Ilicic, T., Henriksson, J., Natarajan, K. N., Tuck, A. C., Gao, X., Bühler, M., Liu, P., Marioni, J. C., and Teichmann, S. A. (2015). Single cell rna-sequencing of pluripotent states unlocks modular transcriptional variation. *Cell Stem Cell*, 17(4):471 – 485.

Kunath, T., Saba-El-Leil, M. K., Almousailleakh, M., , J., Meloche, S., and Smith, A. (2007). Fgf stimulation of the erk1/2 signalling cascade triggers transition of pluripotent embryonic stem cells from self-renewal to lineage commitment. *Development*, 134(16):2895–2902.

Kuroda, T., Tada, M., Kubota, H., Kimura, H., Hatano, S.-y., Suemori, H., Nakatsuji, N., and Tada, T. (2005). Octamer and sox elements are required for transcriptional cis regulation of nanog gene expression. *Molecular and cellular biology*, 25(6):2475–2485.

Kursawe, J., Bardenet, R., Zartman, J. J., Baker, R. E., and Fletcher, A. G. (2016). Robust cell tracking in epithelial tissues through identification of maximum common subgraphs. *Journal of The Royal Society Interface*, 13(124):20160725.

Lanner, F. and Rossant, J. (2010). The role of fgf/erk signaling in pluripotent cells. *Development*, 137(20):3351–3360.

Lechler, T. and Fuchs, E. (2005). Asymmetric cell divisions promote stratification and differentiation of mammalian skin. *Nature*, 437(7056):275.

Leeb, M., Dietmann, S., Paramor, M., Niwa, H., and Smith, A. (2014). Genetic exploration of the exit from self-renewal using haploid embryonic stem cells. *Cell stem cell*, 14(3):385–393.

Li, Z., Fei, T., Zhang, J., Zhu, G., Wang, L., Lu, D., Chi, X., Teng, Y., Hou, N., Yang, X., et al. (2012). Bmp4 signaling acts via dual-specificity phosphatase 9 to control erk activity in mouse embryonic stem cells. *Cell Stem Cell*, 10(2):171–182.

Lodish, H. (2008). *Molecular cell biology*. Macmillan.

Loh, Y.-H., Wu, Q., Chew, J.-L., Vega, V. B., Zhang, W., Chen, X., Bourque, G., George, J., Leong, B., Liu, J., et al. (2006). The oct4 and nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nature genetics*, 38(4):431–440.

Lowell, S., Benchoua, A., Heavey, B., and Smith, A. G. (2006). Notch promotes neural lineage entry by pluripotent embryonic stem cells. *PLoS biology*, 4(5):e121.

MacArthur, R. H. (1957). On the relative abundance of bird species. *Proceedings of the National Academy of Sciences*, 43(3):293–295.

Marks, H., Kalkan, T., Menafra, R., Denissov, S., Jones, K., Hofemeister, H., Nichols, J., Kranz, A., Stewart, A. F., Smith, A., et al. (2012). The transcriptional and epigenomic foundations of ground state pluripotency. *Cell*, 149(3):590–604.

Martello, G., Bertone, P., and Smith, A. (2013). Identification of the missing pluripotency mediator downstream of leukaemia inhibitory factor. *The EMBO journal*, 32(19):2561–2574.

Martello, G. and Smith, A. (2014). The Nature of Embryonic Stem Cells. *Annual Review of Cell and Developmental Biology*, 30(1):647–675.

Martello, G., Sugimoto, T., Diamanti, E., Joshi, A., Hannah, R., Ohtsuka, S., Göttgens, B., Niwa, H., and Smith, A. (2012). Esrrb is a pivotal target of the gsk3/tcf3 axis regulating embryonic stem cell self-renewal. *Cell stem cell*, 11(4):491–504.

Martin, G. R. (1980). Teratocarcinomas and mammalian embryogenesis. *Science*, 209(4458):768–776.

Martin, G. R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proceedings of the National Academy of Sciences*, 78(12):7634–7638.

Martin, G. R. and Evans, M. J. (1975). The formation of embryoid bodies in vitro by homogeneous embryonal carcinoma cell cultures derived from isolated single cells. *Teratomas and differentiation*, pages 169–187.

Martin, G. R., Wiley, L. M., and Damjanov, I. (1977). The development of cystic embryoid bodies in vitro from clonal teratocarcinoma stem cells. *Developmental biology*, 61(2):230–244.

Maruyama, M., Ichisaka, T., Nakagawa, M., and Yamanaka, S. (2005). Differential roles for sox15 and sox2 in transcriptional control in mouse embryonic stem cells. *Journal of Biological Chemistry*, 280(26):24371–24379.

Masui, S., Nakatake, Y., Toyooka, Y., Shimosato, D., Yagi, R., Takahashi, K., Okochi, H., Okuda, A., Matoba, R., Sharov, A. A., et al. (2007). Pluripotency governed by sox2 via regulation of oct3/4 expression in mouse embryonic stem cells. *Nature cell biology*, 9(6):625–635.

Masui, S., Ohtsuka, S., Yagi, R., Takahashi, K., Ko, M. S., and Niwa, H. (2008). Rex1/zfp42 is dispensable for pluripotency in mouse es cells. *BMC developmental biology*, 8(1):45.

Matsuda, T., Nakamura, T., Nakao, K., Arai, T., Katsuki, M., Heike, T., and Yokota, T. (1999). Stat3 activation is sufficient to maintain an undifferentiated state of mouse embryonic stem cells. *The EMBO journal*, 18(15):4261–4269.

Mazumdar, J., O'brien, W. T., Johnson, R. S., LaManna, J. C., Chavez, J. C., Klein, P. S., and Simon, M. C. (2010). O2 regulates stem cells through wnt/[beta]-catenin signalling. *Nature cell biology*, 12(10):1007–1013.

Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M., and Yamanaka, S. (2003). The homeoprotein nanog is required for maintenance of pluripotency in mouse epiblast and es cells. *cell*, 113(5):631–642.

Moris, N., Pina, C., and Arias, A. M. (2016). Transition states and cell fate decisions in epigenetic landscapes. *Nature Reviews Genetics*, 17(11):693–703.

Morrison, S. J. and Kimble, J. (2006). Asymmetric and symmetric stem-cell divisions in development and cancer. *nature*, 441(7097):1068.

Mulas, C., Kalkan, T., and Smith, A. (2017). Nodal secures pluripotency upon embryonic stem cell progression from the ground state. *Stem Cell Reports*, 9(1):77–91.

Muzzey, D. and van Oudenaarden, A. (2009). Quantitative time-lapse fluorescence microscopy in single cells. *Annual Review of Cell and Developmental*, 25:301–327.

Nakai-Futatsugi, Y. and Niwa, H. (2016). Zscan4 is activated after telomere shortening in mouse embryonic stem cells. *Stem cell reports*, 6(4):483–495.

Ng, H.-H. and Surani, M. A. (2011). The transcriptional and signalling networks of pluripotency. *Nature Cell Biology*, 13(5):490–496.

Nichols, J. and Smith, A. (2009). Naive and primed pluripotent states. *Cell stem cell*, 4(6):487–492.

Nichols, J. and Smith, A. (2012). Pluripotency in the embryo and in culture. *Cold Spring Harbor perspectives in biology*, 4(8):a008128.

Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Schöler, H., and Smith, A. (1998). Formation of pluripotent stem cells in the mammalian embryo depends on the pou transcription factor oct4. *Cell*, 95(3):379–391.

Niwa, H., Burdon, T., Chambers, I., and Smith, A. (1998). Self-renewal of pluripotent embryonic stem cells is mediated via activation of stat3. *Genes & development*, 12(13):2048–2060.

Niwa, H., Miyazaki, J.-i., and Smith, A. G. (2000). Quantitative expression of oct-3/4 defines differentiation, dedifferentiation or self-renewal of es cells. *Nature genetics*, 24(4):372–376.

Niwa, H., Ogawa, K., Shimosato, D., and Adachi, K. (2009). A parallel circuit of lif signalling pathways maintains pluripotency of mouse es cells. *Nature*, 460(7251):118–122.

Ohnishi, Y., Huber, W., Tsumura, A., Kang, M., Xenopoulos, P., Kurimoto, K., Oleś, A. K., Araúzo-Bravo, M. J., Saitou, M., Hadjantonakis, A.-K., et al. (2013). Cell-to-cell expression variability followed by signal reinforcement progressively segregates early mouse lineages. *Nature cell biology*, 16(1):ncb2881.

Okita, K., Ichisaka, T., and Yamanaka, S. (2007). Generation of germline-competent induced pluripotent stem cells. *nature*, 448(7151):313–317.

Padovan-Merhar, O., Nair, G. P., Biaesch, A. G., Mayer, A., Scarfone, S., Foley, S. W., Wu, A. R., Churchman, L. S., Singh, A., and Raj, A. (2015). Single mammalian cells compensate for differences in cellular volume and dna copy number through independent global transcriptional mechanisms. *Molecular cell*, 58(2):339–352.

Pauklin, S. and Vallier, L. (2013). The cell-cycle state of stem cells determines cell fate propensity. *Cell*, 155(1):135–147.

Pauklin, S. and Vallier, L. (2014). The cell-cycle state of stem cells determines cell fate propensity. *Cell*, 156(6):1338.

Pelton, T., Sharma, S., Schulz, T., Rathjen, J., and Rathjen, P. D. (2002). Transient pluripotent cell populations during primitive ectoderm formation: correlation of in vivo and in vitro pluripotent cell development. *Journal of cell science*, 115(2):329–339.

Pereira, L., Yi, F., and Merrill, B. J. (2006). Repression of nanog gene transcription by tcf3 limits embryonic stem cell self-renewal. *Molecular and cellular biology*, 26(20):7479–7491.

Pesce, M. and Schöler, H. R. (2001). Oct-4: gatekeeper in the beginnings of mammalian development. *Stem cells*, 19(4):271–278.

Quinlan, L. (2011). Signaling pathways in mouse embryo stem cell self-renewal. In *Embryonic Stem Cells-Basic Biology to Bioengineering*. InTech.

Rasband, W., ImageJ, and U.S. National Institute of Health, Bethesda, Maryland, USA (1997–2014). http://imagej.nih.gov/ij/.

Reynolds, N., Latos, P., Hynes-Allen, A., Loos, R., Leaford, D., O'Shaughnessy, A., Mosaku, O., Signolet, J., Brennecke, P., Kalkan, T., et al. (2012). Nurd suppresses pluripotency gene expression to promote transcriptional heterogeneity and lineage commitment. *Cell stem cell*, 10(5):583–594.

Rezende, N. C., Lee, M.-Y., Monette, S., Mark, W., Lu, A., and Gudas, L. J. (2011). Rex1 (zfp42) null mice show impaired testicular function, abnormal testis morphology, and aberrant gene expression. *Developmental biology*, 356(2):370–382.

Rodda, D. J., Chew, J.-L., Lim, L.-H., Loh, Y.-H., Wang, B., Ng, H.-H., and Robson, P. (2005). Transcriptional regulation of nanog by oct4 and sox2. *Journal of Biological Chemistry*, 280(26):24731–24737.

Rose, L. and Gönczy, P. (2014). Polarity establishment, asymmetric division and segregation of fate determinants in early c. elegans embryos. *WormBook*.

Rosin, P. (1933). The laws governing the fineness of powdered coal. *J. Inst. Fuel.*, 7:29–36.

Ruzinova, M. B. and Benezra, R. (2003). Id proteins in development, cell cycle and cancer. *Trends in cell biology*, 13(8):410–418.

Saiz, N., Williams, K. M., Seshan, V. E., and Hadjantonakis, A.-K. (2016). Asynchronous fate decisions by single cells collectively ensure consistent lineage composition in the mouse blastocyst. *Nature communications*, 7:13463.

Sakaue-Sawano, A., Kurokawa, H., Morimura, T., Hanyu, A., Hama, H., Osawa, H., Kashiwagi, S., Fukami, K., Miyata, T., Miyoshi, H., et al. (2008a). Visualizing spatiotemporal dynamics of multicellular cell-cycle progression. *Cell*, 132(3):487–498.

Sakaue-Sawano, A., Ohtawa, K., Hama, H., Kawano, M., Ogawa, M., and Miyawaki, A. (2008b). Tracing the silhouette of individual cells in s/g 2/m phases with fluorescence. *Chemistry & biology*, 15(12):1243–1248.

Sandler, O., Mizrahi, S. P., Weiss, N., Agam, O., Simon, I., and Balaban, N. Q. (2015). Lineage correlations of single cell division time as a probe of cell-cycle dynamics. *Nature*, 519(7544):468–483.

Scarpa, E. and Mayor, R. (2016). Collective cell migration in development. *J Cell Biol*, 212(2):143–155.

Scherf, N., Franke, K., Glauche, I., Kurth, I., Bornhäuser, M., Werner, C., Pompe, T., and Roeder, I. (2012). On the symmetry of siblings: automated single-cell tracking to quantify the behavior of hematopoietic stem cells in a biomimetic setup. *Experimental hematology*, 40(2):119–130.

Schöler, H. R., Ruppert, S., Suzuki, N., Chowdhury, K., and Gruss, P. (1990). New type of pou domain in germ line-specific protein oct-4. *Nature*, 344(6265):435–439.

Schroeder, T. (2011). Long-term single-cell imaging of mammalian stem cells. *Nature methods*, 8(4s):S30–S35.

Shahbazi, M. N., Scialdone, A., Skorupska, N., Weberling, A., Recher, G., Zhu, M., Jedrusik, A., Devito, L. G., Noli, L., Macaulay, I. C., et al. (2017). Pluripotent state transitions coordinate morphogenesis in mouse and human embryos. *Nature*, 552(7684):239.

Shi, W., Wang, H., Pan, G., Geng, Y., Guo, Y., and Pei, D. (2006). Regulation of the pluripotency marker Rex-1 by Nanog and Sox2. *Journal of Biological Chemistry*, 281(33):23319–23325.

Silva, J., Nichols, J., Theunissen, T. W., Guo, G., van Oosten, A. L., Barrandon, O., Wray, J., Yamanaka, S., Chambers, I., and Smith, A. (2009). Nanog is the gateway to the pluripotent ground state. *Cell*, 138(4):722–737.

Simons, B. D. and Clevers, H. (2011). Strategies for homeostatic stem cell self-renewal in adult tissues. *Cell*, 145(6):851–862.

Skylaki, S., Hilsenbeck, O., and Schroeder, T. (2016). Challenges in long-term imaging and quantification of single-cell dynamics. *Nature biotechnology*, 34(11):1137–1144.

Smith, A. (1998). Cell therapy: in search of pluripotency. *Current biology*, 8(22):R802–R804.

Smith, A. (2006). A glossary for stem-cell biology. *Nature*, 441(7097):1060–1060.

Smith, A. (2017). Formative pluripotency: the executive phase in a developmental continuum. *Development*, 144(3):365–373.

Smith, A. G. (2001). Embryo-derived stem cells: of mice and men. *Annual review of cell and developmental biology*, 17(1):435–462.

Smith, A. G., Heath, J. K., Donaldson, D. D., Wong, G. G., Moreau, J., Stahl, M., and Roger, D. (1988). Inhibition of stem cell differentiation by purified polypeptides. *Nature*, 336:15.

Smith, A. G. and Hooper, M. L. (1987). Buffalo rat liver cells produce a diffusible activity which inhibits the differentiation of murine embryonal carcinoma and embryonic stem cells. *Developmental biology*, 121(1):1–9.

Snow, M. H. L. (1977). Gastrulation in the mouse: Growth and regionalization of the epiblast. *Development*, 42(1):293–303.

Solter, D., ŠKREB, N., and DAMJANOV, I. (1970). Extrauterine growth of mouse egg-cylinders results in malignant teratoma. *Nature*, 227(5257):503–504.

Song, X., Zhu, C.-H., Doan, C., and Xie, T. (2002). Germline stem cells anchored by adherens junctions in the drosophila ovary niches. *Science*, 296(5574):1855–1857.

Stauffer, H. (1979). A derivation for the weibull distribution. *Journal of theoretical biology*, 81(1):55–63.

Stevens, L. C. (1958). Studies on transplantable testicular teratomas of strain 129 mice. *JNCI: Journal of the National Cancer Institute*, 20(6):1257–1275.

Stevens, L. C. (1959). Embryology of testicular teratomas in strain 129 mice. *Journal of the National Cancer Institute*, 23(6):1249–1295.

Stevens, L. C. (1970). The development of transplantable teratocarcinomas from intratesticular grafts of pre-and postimplantation mouse embryos. *Developmental biology*, 21(3):364–382.

Stevens, L. C. and Little, C. C. (1954). Spontaneous testicular teratomas in an inbred strain of mice. *Proceedings of the National Academy of Sciences*, 40(11):1080–1087.

Tajbakhsh, S., Rocheteau, P., and Le Roux, I. (2009). Asymmetric cell divisions and asymmetric cell fates. *Annual Review of Cell and Developmental*, 25:671–699.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *cell*, 131(5):861–872.

Takahashi, K. and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *cell*, 126(4):663–676.

Tanimura, S. and Takeda, K. (2017). Erk signalling as a regulator of cell motility. *The Journal of Biochemistry*, 162(3):145–154.

ten Berge, D., Kurek, D., Blauwkamp, T., Koole, W., Maas, A., Eroglu, E., Siu, R. K., and Nusse, R. (2011). Embryonic stem cells require wnt proteins to prevent differentiation to epiblast stem cells. *Nature cell biology*, 13(9):1070–1075.

ter Huurne, M., Chappell, J., Dalton, S., and Stunnenberg, H. G. (2017). Distinct cell-cycle control in two different states of mouse pluripotency. *Cell Stem Cell*, 21(4):449–455.

Tian, X., Liu, Z., Niu, B., Zhang, J., Tan, T. K., Lee, S. R., Zhao, Y., Harris, D. C., and Zheng, G. (2011). E-cadherin/β-catenin complex and the epithelial barrier. *BioMed Research International*, 2011.

Torres-Padilla, M.-E. and Chambers, I. (2014). Transcription factor heterogeneity in pluripotent stem cells: a stochastic advantage. *Development*, 141(11):2173–2181.

Toyooka, Y., Shimosato, D., Murakami, K., Takahashi, K., and Niwa, H. (2008). Identification and characterization of subpopulations in undifferentiated es cell culture. *Development*, 135(5):909–918.

Ulman, V., Maška, M., Magnusson, K. E., Ronneberger, O., Haubold, C., Harder, N., Matula, P., Matula, P., Svoboda, D., Radojevic, M., et al. (2017). An objective comparison of cell-tracking algorithms. *Nature methods*, 14:1141.

Waisman, A., Vazquez, E. C., Solari, C., Cosentino, M. S., Martyn, I., Deglincerti, A., Ozair, M. Z., Ruzo, A., Barañao, L., Miriuka, S., et al. (2017). Inhibition of cell division and dna replication impair mouse-naïve pluripotency exit. *Journal of molecular biology*, 429(18):2802.

Weibull, W. et al. (1951). A statistical distribution function of wide applicability. *Journal of applied mechanics*, 18(3):293–297.

Williams, R. L., Hilton, D. J., Pease, S., Willson, T. A., Stewart, C. L., Gearing, D. P., Wagner, E. F., Metcalf, D., Nicola, N. A., and Gough, N. M. (1988). Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature*, 336(6200):684–687.

Winter, M., Wait, E., Roysam, B., Goderie, S. K., Ali, R. A. N., Kokovay, E., Temple, S., and Cohen, A. R. (2011). Vertebrate neural stem cell segmentation, tracking and lineaging with validation and editing. *Nature Protocols*, 6(12):1942–1952.

Wray, J., Kalkan, T., Gomez-Lopez, S., Eckardt, D., Cook, A., Kemler, R., and Smith, A. (2011). Inhibition of glycogen synthase kinase-3 alleviates tcf3 repression of the pluripotency network and increases embryonic stem cell resistance to differentiation. *Nature cell biology*, 13(7):838–845.

Wray, J., Kalkan, T., and Smith, A. (2010). The ground state of pluripotency. *Biochemical Society Transactions*, 38(4):1027–1032.

Wu, D. and Pan, W. (2010). Gsk3: a multifaceted kinase in wnt signaling. *Trends in biochemical sciences*, 35(3):161–168.

Yamaguchi, S., Kimura, H., Tada, M., Nakatsuji, N., and Tada, T. (2005). Nanog expression in mouse germ cell development. *Gene Expression Patterns*, 5(5):639–646.

Yamanaka, Y., Lanner, F., and Rossant, J. (2010). Fgf signal-dependent segregation of primitive endoderm and epiblast in the mouse blastocyst. *Development*, 137(5):715–724.

Yamashita, Y. M., Fuller, M. T., and Jones, D. L. (2005). Signaling in stem cell niches: lessons from the drosophila germline. *Journal of cell science*, 118(4):665–672.

Yamashita, Y. M., Jones, D. L., and Fuller, M. T. (2003). Orientation of asymmetric stem cell division by the apc tumor suppressor and centrosome. *Science*, 301(5639):1547–1550.

Yeo, J.-C., Jiang, J., Tan, Z.-Y., Yim, G.-R., Ng, J.-H., Göke, J., Kraus, P., Liang, H., Gonzales, K. A. U., Chong, H.-C., et al. (2014). Klf2 is an essential factor that sustains ground state pluripotency. *Cell Stem Cell*, 14(6):864–872.

Yi, F., Pereira, L., and Merrill, B. J. (2008). Tcf3 functions as a steady-state limiter of transcriptional programs of mouse embryonic stem cell self-renewal. *Stem Cells*, 26(8):1951–1960.

Ying, Q.-L., Nichols, J., Chambers, I., and Smith, A. (2003). Bmp induction of id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with stat3. *Cell*, 115(3):281–292.

Ying, Q.-L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. *Nature*, 453(7194):519–523.

Yordanov, B., Dunn, S.-J., Kugler, H., Smith, A., Martello, G., and Emmott, S. (2016). A method to identify and analyze biological programs through automated reasoning. *NPJ systems biology and applications*, 2:16010.

Young, R. A. (2011). Control of the embryonic stem cell state. *Cell*, 144(6):940–954.

Zhong, W., Jiang, M.-M., Weinmaster, G., Jan, L. Y., and Jan, Y. N. (1997). Differential expression of mammalian numb, numblike and notch1 suggests distinct roles during mouse cortical neurogenesis. *Development*, 124(10):1887–1897.

Zhou, X., Smith, A. J., Waterhouse, A., Blin, G., Malaguti, M., Lin, C.-Y., Osorno, R., Chambers, I., and Lowell, S. (2013). Hes1 desynchronizes differentiation of pluripotent cells by modulating stat3 activity. *Stem Cells*, 31(8):1511–1522.

Appendix A

Kolmogorov-Smirnov Tests for Normality of Generational Attribute Distributions

Cells were sorted into distributions by generation to track the change in cell-cycle duration, cell motility, and cell area over time. ANOVA testing with Holms-Sidak post test was used to compare within conditions and between conditions. Prior to ANOVA, to ensure the validity of the test, the generational distributions for 2i ES cells and N2B27 Cells for both replicates were tested for normality using MATLAB's Kolmogorov-Smirnov Test with a 99.9% confidence interval.

To compare the empirical distributions to the normal distribution, with zero mean and unit variance ($\mathcal{N}(0,1)$), each empirical distribution was transformed by subtracting by the mean, and then dividing by the standard deviation. If the data is normal, then the transformed empirical distribution should be equal to $\mathcal{N}(0,1)$. Each set of tests are show with a table that shows the the binary accept/reject value for the null hypothesis (H0), the p-value, and the Kolmogorow-Smirnov (K-S) Statistic. The null-hypothesis is that the transformed empirical distribution is normally distributed with zero mean and unit variance. If the empirical distribution is determined to be $\mathcal{N}(0,1)$, then H0 = 0, otherwise H0 = 1.

All generations for both replicates of 2i ES cells and N2B27 Cells are normally distributed except for one. Only generation 3 N2B27 cells of replicate 1 for cell-cycle duration was found to be non-normal with a p-value = 5×10^5. However, given that all other generations for cell-cycle duration of both replicates in both conditions are normally distributed, it is sufficient to say that an assumption of normality can be made.

A.1 Mean Cell Area

	Mean Cell Area (99.9% CI)					
Generation	2i			N2B27		
	H0	p-value	K-S Stat	H0	p-value	K-S Stat
1	0	0.89	0.10	0	0.92	0.05
	0	0.97	0.10	0	0.59	0.10
2	0	0.36	0.12	0	0.28	0.07
	0	0.58	0.16	0	0.26	0.10
3	0	0.20	0.14	0	0.23	0.08
	0	0.52	0.24	0	0.50	0.10
4	0	0.94	0.15	0	0.17	0.11
	-	-	-	0	0.36	0.24
5	-	-	-	0	0.56	0.16
	-	-	-	-	-	-

(Null hypothesis, the empirical distribution is normal. 0 accept Null, 1 reject null.)

Fig. A.1 The table of results from the Kolmogorov-Smirnov Test for each generational distribution of Mean Cell Area. A 99.9% confidence interval was used in testing each generation of 2i ES cells and N2B27 cells for both replicates, which are housed in each row. All of the distributions are normally distributed.

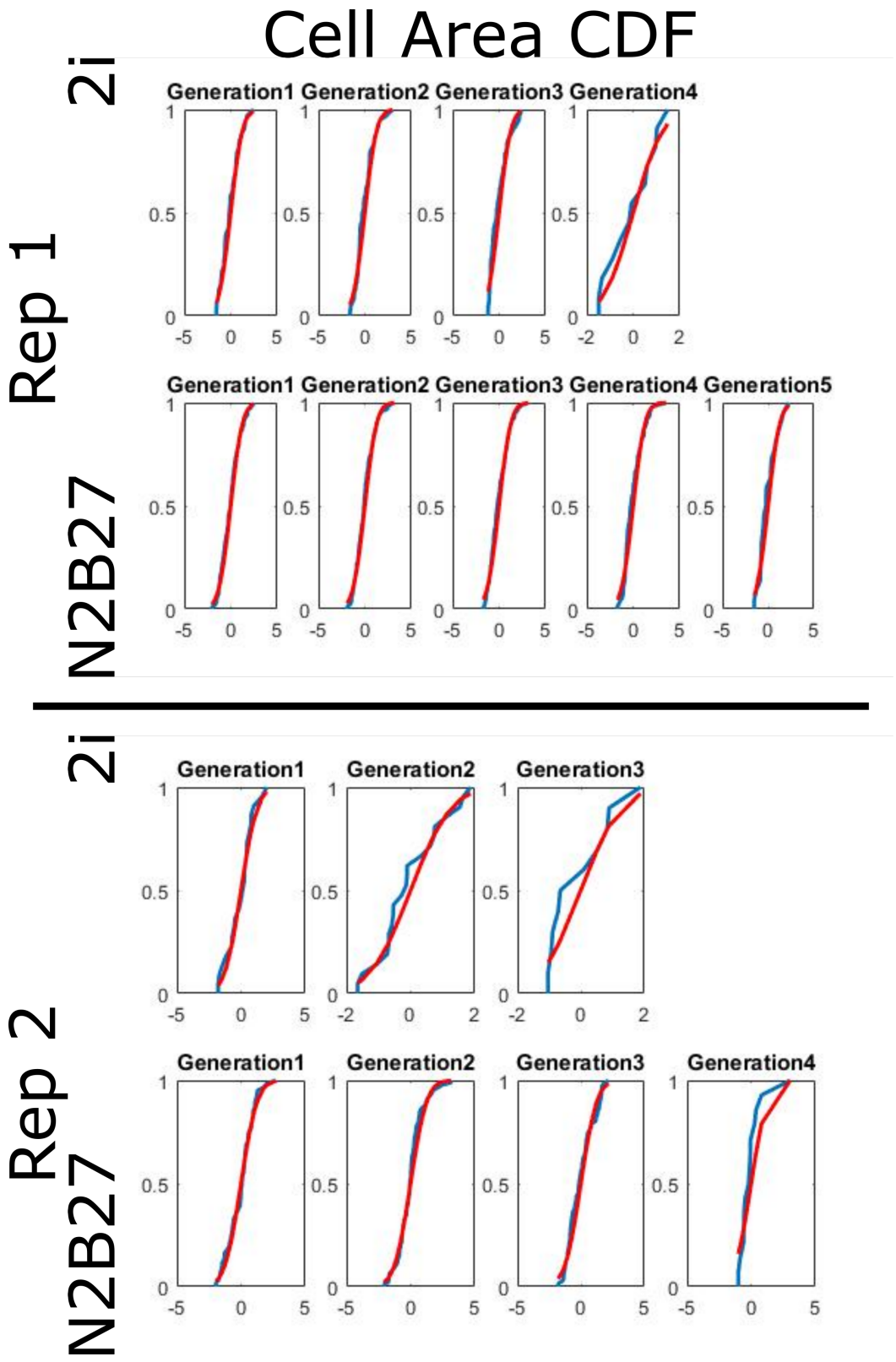


Fig. A.2 Cumulative Distribution functions for the Kolmogorov-Smirnov Test. The Red curve is the CDF of $\mathcal{N}(0,1)$ and the Blue curve is the transformed empirical CDF of mean cell area. Replicate one is above the central black line and replicate 2 is below the central black line. For each replicate the generational distributions for 2i ES cells and N2B27 cells are shown in rows, starting with generation 1 at the left.

A.2 Cell-Cycle Duration

	Cell Cycle (99.9% CI)					
Generation	2i			N2B27		
	H0	p-value	K-S Stat	H0	p-value	K-S Stat
1	0	0.32	0.16	0	0.05	0.12
	0	0.66	0.15	0	0.15	0.14
2	0	0.30	0.12	0	0.03	0.11
	0	0.98	0.094	0	0.03	0.15
3	0	0.55	0.10	**1**	**5e-4**	**0.15**
	0	0.56	0.23	0	0.27	0.12
4	0	0.44	0.25	0	0.02	0.15
	-	-	-	0	0.39	0.23
5	-	-	-	0	0.44	0.18
	-	-	-	-	-	-

(Null hypothesis, the empirical distribution is normal. 0 accept Null, 1 reject null.)

Fig. A.3 The table of results from the Kolmogorov-Smirnov Test for each generational distribution of cell-cycle duration. A 99.9% confidence interval was used in testing each generation of 2i ES cells and N2B27 cells for both replicates, which are housed in each row. Only generation 3 N2B27 cells of replicate 1 are found to be non-normal, However, given that all other generations for both replicates in both conditions are normal, it is sufficient to say that an assumption of normality can be made.

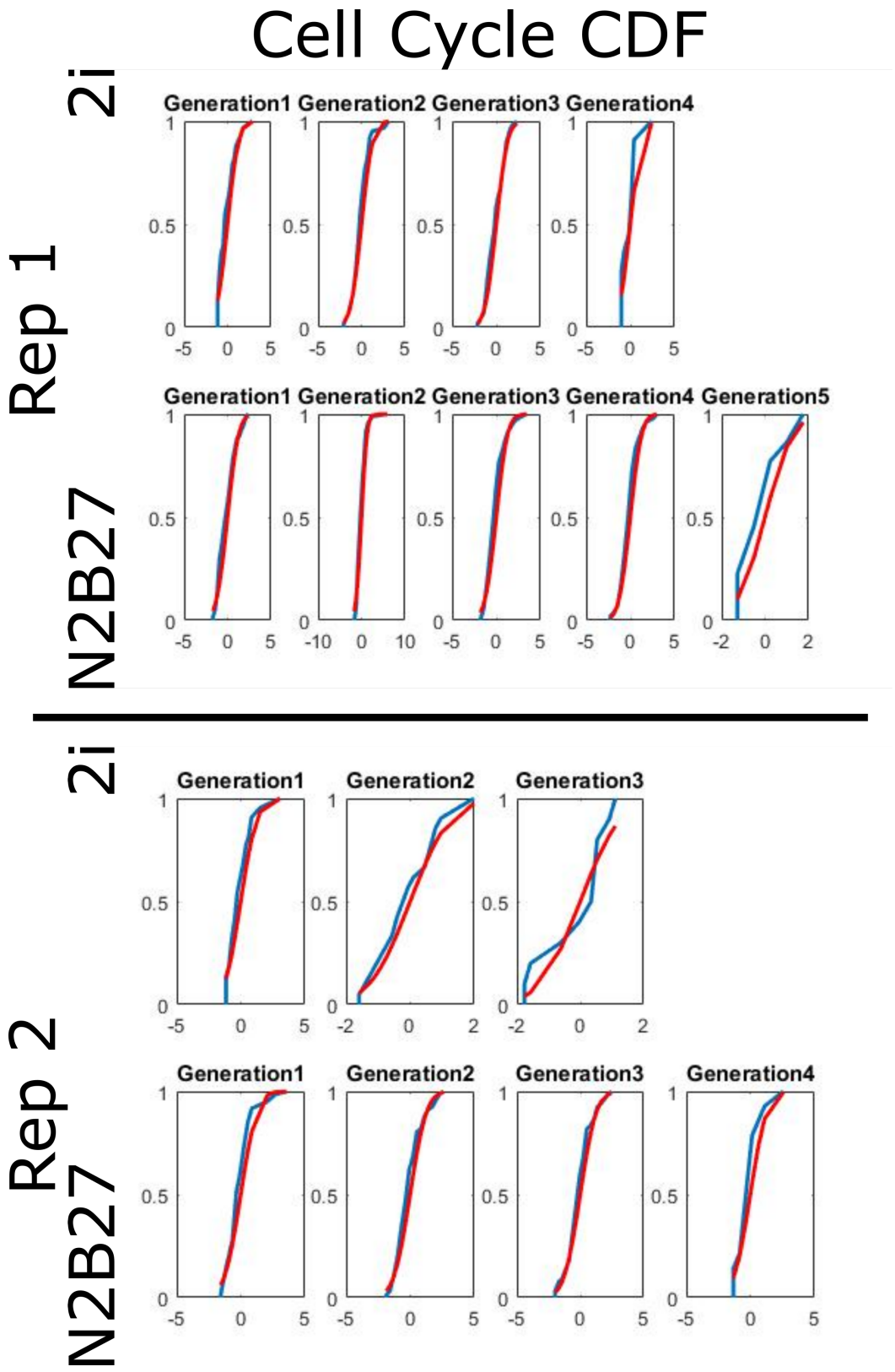


Fig. A.4 Cumulative Distribution functions for the Kolmogorov-Smirnov Test. The Red curve is the CDF of $\mathcal{N}(0,1)$ and the Blue curve is the transformed empirical CDF of cell-cycle duration. Replicate one is above the central black line and replicate 2 is below the central black line. For each replicate the generational distributions for 2i ES cells and N2B27 cells are shown in rows, starting with generation 1 at the left.

A.3 Motility

	Motility (99.9% CI)					
Generation	2i			N2B27		
	H0	p-value	K-S Stat	H0	p-value	K-S Stat
1	0	0.93	0.09	0	0.48	0.07
	0	0.70	0.14	0	0.28	0.12
2	0	0.59	0.10	0	0.14	0.09
	0	0.90	0.12	0	0.97	0.05
3	0	0.79	0.08	0	0.94	0.04
	0	0.67	0.21	0	0.98	0.05
4	0	0.71	0.20	0	0.89	0.06
	-	-	-	0	0.93	0.13
5	-	-	-	0	0.20	0.33
	-	-	-	-	-	-

(Null hypothesis, the empirical distribution is normal. 0 accept Null, 1 reject null.)

Fig. A.5 The table of results from the Kolmogorov-Smirnov Test for each generational distribution of Motility. A 99.9% confidence interval was used in testing each generation of 2i ES cells and N2B27 cells for both replicates, which are housed in each row. All of the distributions are normally distributed.

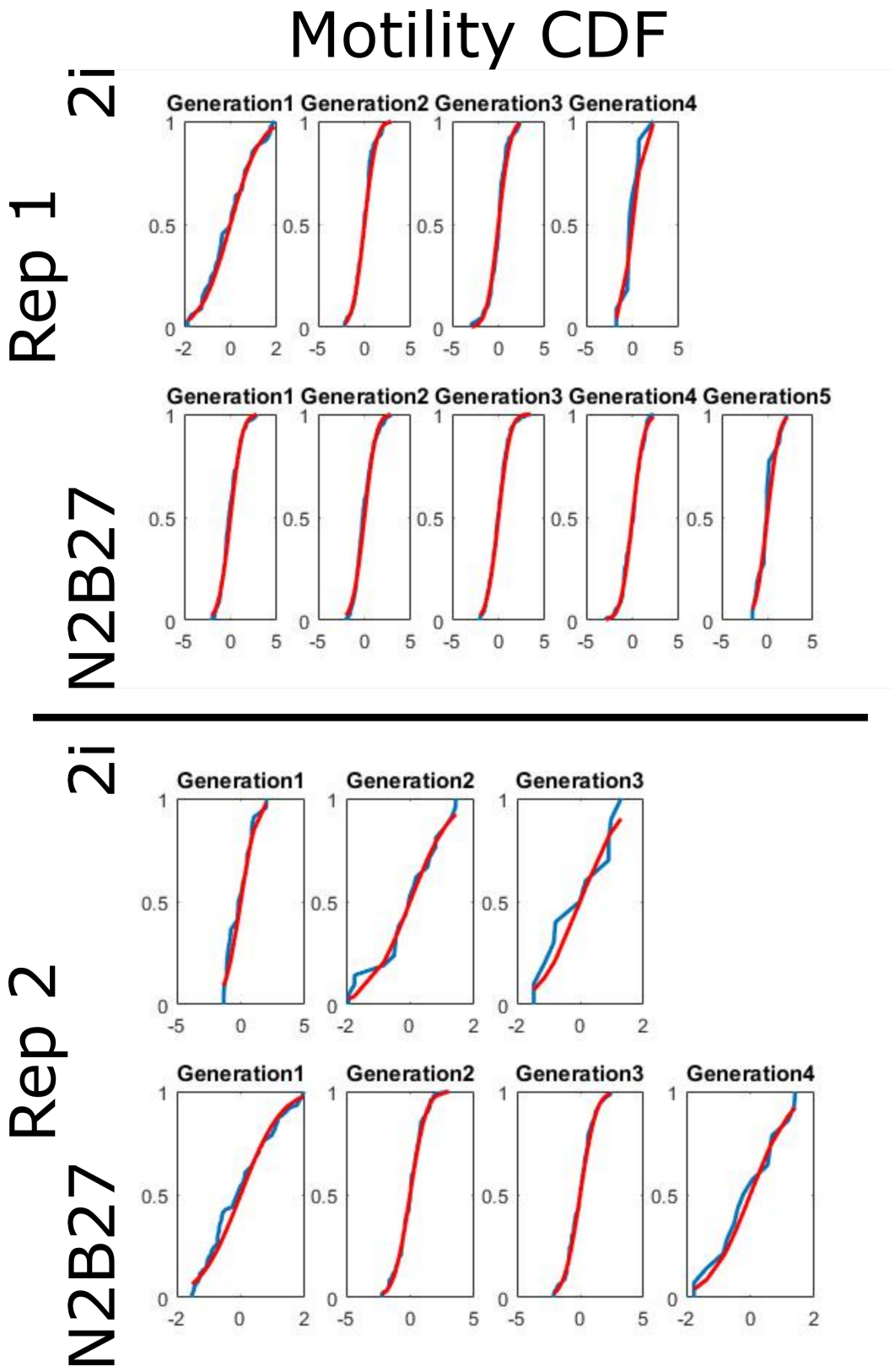


Fig. A.6 Cumulative Distribution functions for the Kolmogorov-Smirnov Test. The Red curve is the CDF of $\mathcal{N}(0,1)$ and the Blue curve is the transformed empirical CDF of motility. Replicate one is above the central black line and replicate 2 is below the central black line. For each replicate the generational distributions for 2i ES cells and N2B27 cells are shown in rows, starting with generation 1 at the left.